"Express Mail" Mailing Label No. EL 341 578 030 US

Atty. Docket No. 2300-1492

Date of Deposit September 23, 1999

I certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to Box Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Signature: Bonnie F. Wharton

Typed Name of Person Mailing Paper or Fee: Bonnie F. Wharton

UTILITY APPLICATION

FOR

HUMAN GENES AND GENE EXPRESSION PRODUCTS

Carol L. Francis
Registration No. 36,513
BOZICEVIC, FIELD & FRANCIS LLP
285 Hamilton Avenue, Suite 200
Palo Alto, CA 94301
Telephone: (650) 327-3400
Facsimile: (650) 327-3231

F:\MJO\Docs\Pats\covers\conversion cover.frm

HUMAN GENES AND GENE EXPRESSION PRODUCTS

Cross-Reference to Related Applications

This application claims the benefit of U.S. provisional patent application serial no. 60/102,180, filed September 28, 1998; and U.S. provisional patent application serial no. 60/102,161, filed September 28, 1998; and U.S. provisional patent application serial no. 60/102,380, filed September 29, 1998; and U.S. provisional patent application serial no. 60/103,815, filed October 8, 1998; and U.S. provisional patent application serial no. 60/105,877 filed October 27, 1998; each of which applications are incorporated herein by reference.

Field of the Invention

The present invention relates to polynucleotides of human origin and the encoded gene products.

Background of the Invention

Identification of novel polynucleotides, particularly those that encode an expressed gene product, is important in the advancement of drug discovery, diagnostic technologies, and the understanding of the progression and nature of complex diseases such as cancer. Identification of genes expressed in different cell types isolated from sources that differ in disease state or stage, developmental stage, exposure to various environmental factors, the tissue of origin, the species from which the tissue was isolated, and the like is key to identifying the genetic factors that are responsible for the phenotypes associated with these various differences.

This invention provides novel human polynucleotides, the polypeptides encoded by these polynucleotides, and the genes and proteins corresponding to these novel polynucleotides.

Summary of the Invention

This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides and to proteins expressed by the genes. The invention also relates to diagnostics and therapeutics comprising such novel human polynucleotides, their corresponding genes or gene products, including probes, antisense nucleotides, and antibodies. The polynucleotides of the invention correspond to a polynucleotide comprising the sequence information of at least one of SEQ ID NOS:1-1079.

Various aspects and embodiments of the invention will be readily apparent to the ordinarily skilled artisan upon reading the description provided herein.

Detailed Description of the Invention

The invention relates to polynucleotides comprising the disclosed nucleotide sequences, to full length cDNA, mRNA genomic sequences, and genes corresponding to these sequences and degenerate variants thereof, and to polypeptides encoded by the polynucleotides of the invention

and polypeptide variants. The following detailed description describes the polynucleotide compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these polynucleotides and genes, identification of structural motifs of the polynucleotides and genes, identification of the function of a gene product encoded by a gene corresponding to a polynucleotide of the invention, use of the provided polynucleotides as probes and in mapping and in tissue profiling, use of the corresponding polypeptides and other gene products to raise antibodies, and use of the polynucleotides and their encoded gene products for therapeutic and diagnostic purposes.

Polynucleotide Compositions

The scope of the invention with respect to polynucleotide compositions includes, but is not necessarily limited to, polynucleotides having a sequence set forth in any one of SEQ ID NOS:1-1079; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here. "Polynucleotide" and "nucleic acid" as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicted.

The invention features polynucleotides that are expressed in human tissue, specifically human colon, breast, and/or lung tissue. Novel nucleic acid compositions of the invention of particular interest comprise a sequence set forth in any one of SEQ ID NOS:1-1079 or an identifying sequence thereof. An "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a polynucleotide sequence, *e.g.*, exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of SEQ ID NOS: 1-1079.

The polynucleotides of the invention also include polynucleotides having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by

hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, USPN 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences (SEQ ID NOS:1-1079) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, *e.g.* primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

Preferably, hybridization is performed using at least 15 contiguous nucleotides (nt) of at least one of SEQ ID NOS:1-1079. That is, when at least 15 contiguous nt of one of the disclosed SEQ ID NOS. is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one SEQ ID NO. can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA. Probes of more than 15 nt can be used, e.g., probes of from about 18 nt to about 100 nt, but 15 nt represents sufficient sequence for unique identification.

The polynucleotides of the invention also include naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants, allelic variants, *etc.*). Variants of the polynucleotides of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

The invention also encompasses homologs corresponding to the polynucleotides of SEQ ID NOS:1-1079, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, etc. Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 contiguous nt

long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as gapped BLAST, described in Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90% or more as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (*e.g.*, in diagnosis, as a unique identifier of a differentially expressed gene of interest, *etc.*). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

The nucleic acid compositions of the subject invention can encode all or a part of the subject polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* Isolated polynucleotides and polynucleotide fragments of the invention comprise at least about 10, about 15, about 20, about 35,

about 50, about 100, about 150 to about 200, about 250 to about 300, or about 350 contiguous nt selected from the polynucleotide sequences as shown in SEQ ID NOS:1-1079. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more. In a preferred embodiment, the polynucleotide molecules comprise a contiguous sequence of at least 12 nt selected from the group consisting of the polynucleotides shown in SEQ ID NOS:1-1079.

Probes specific to the polynucleotides of the invention can be generated using the polynucleotide sequences disclosed in SEQ ID NOS:1-1079. The probes are preferably at least about a 12, 15, 16, 18, 20, 22, 24, or 25 nt fragment of a corresponding contiguous sequence of SEQ ID NOS:1-1079, and can be less than 2, 1, 0.5, 0.1, or 0.05 kb in length. The probes can be synthesized chemically or can be generated from longer polynucleotides using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a polynucleotide of one of SEQ ID NOS:1-1079. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

The polynucleotides of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The polynucleotides of the invention can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides of the invention can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological samples (*e.g.*, extracts of human

cells) to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of the polynucleotide sequences as shown in SEQ ID NOS:1-1079 or variants thereof in a sample. These and other uses are described in more detail below.

Use of Polynucleotides to Obtain Full-Length cDNA, Gene, and Promoter Region

Full-length cDNA molecules comprising the disclosed polynucleotides are obtained as follows. A polynucleotide having a sequence of one of SEQ ID NOS:1-1079, or a portion thereof comprising at least 12, 15, 18, or 20 nt, is used as a hybridization probe to detect hybridizing members of a cDNA library using probe design methods, cloning methods, and clone selection techniques such as those described in USPN 5,654,173. Libraries of cDNA are made from selected tissues, such as normal or tumor tissue, or from tissues of a mammal treated with, for example, a pharmaceutical agent. Preferably, the tissue is the same as the tissue from which the polynucleotides of the invention were isolated, as both the polynucleotides described herein and the cDNA represent expressed genes. Most preferably, the cDNA library is made from the biological material described herein in the Examples. The choice of cell type for library construction can be made after the identity of the protein encoded by the gene corresponding to the polynucleotide of the invention is known. This will indicate which tissue and cell types are likely to express the related gene, and thus represent a suitable source for the mRNA for generating the cDNA. Where the provided polynucleotides are isolated from cDNA libraries, the libraries are prepared from mRNA of human colon cells, more preferably, human colon cancer cells, even more preferably, from a highly metastatic colon cell, Km12L4-A.

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. The cDNA can be prepared by using primers based on sequence from SEQ ID NOS:1-1079. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided polynucleotides, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring

Harbor, NY. In order to obtain additional sequences 5' to the end of a partial cDNA, 5' RACE (*PCR Protocols: A Guide to Methods and Applications*, (1990) Academic Press, Inc.) can be performed.

Genomic DNA is isolated using the provided polynucleotides in a manner similar to the isolation of full-length cDNAs. Briefly, the provided polynucleotides, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the polynucleotides of the invention, but this is not essential. Most preferably, the genomic DNA is obtained from the biological material described herein in the Examples. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook *et al.*, 9.4-9.30. In addition, genomic sequences can be isolated from human BAC libraries, which are commercially available from Research Genetics, Inc., Huntsville, Alabama, USA, for example. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook *et al.*, such that adjacent and overlapping fragments of genomic DNA are isolated. These are mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

Using the polynucleotide sequences of the invention, corresponding full-length genes can be isolated using both classical and PCR methods to construct and probe cDNA libraries. Using either method, Northern blots, preferably, are performed on a number of cell types to determine which cell lines express the gene of interest at the highest level. Classical methods of constructing cDNA libraries are taught in Sambrook *et al.*, *supra*. With these methods, cDNA can be produced from mRNA and inserted into viral or expression vectors. Typically, libraries of mRNA comprising poly(A) tails can be produced with poly(T) primers. Similarly, cDNA libraries can be produced using the instant sequences as primers.

PCR methods are used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant polynucleotides. Such PCR methods include gene trapping and RACE methods. Gene trapping entails inserting a member of a cDNA library into a vector. The vector then is denatured to produce single stranded molecules. Next, a substrate-bound probe, such a biotinylated oligo, is used to trap cDNA inserts of interest. Biotinylated probes can be linked to an avidin-bound solid substrate. PCR methods can be used to amplify the trapped cDNA. To trap sequences corresponding to the full length genes, the labeled probe sequence is based on the polynucleotide sequences of the invention. Random primers or primers specific to the library vector can be used to amplify the trapped cDNA. Such gene trapping techniques are described in Gruber *et al.*, WO 95/04745 and Gruber *et al.*, USPN 5,500,356. Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA.

"Rapid amplification of cDNA ends," or RACE, is a PCR method of amplifying cDNAs from a number of different RNAs. The cDNAs are ligated to an oligonucleotide linker, and amplified by PCR using two primers. One primer is based on sequence from the instant polynucleotides, for which full length sequence is desired, and a second primer comprises sequence that hybridizes to the oligonucleotide linker to amplify the cDNA. A description of this methods is reported in WO 97/19110. In preferred embodiments of RACE, a common primer is designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends (Apte and Siebert, *Biotechniques* (1993) *15*:890-893; Edwards *et al.*, *Nuc. Acids Res.* (1991) *19*:5227-5232). When a single gene-specific RACE primer is paired with the common primer, preferential amplification of sequences between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

Another PCR-based method generates full-length cDNA library with anchored ends without needing specific knowledge of the cDNA sequence. The method uses lock-docking primers (I-VI), where one primer, poly TV (I-III) locks over the polyA tail of eukaryotic mRNA producing first strand synthesis and a second primer, polyGH (IV-VI) locks onto the polyC tail added by terminal deoxynucleotidyl transferase (TdT)(see, e.g., WO 96/40998).

The promoter region of a gene generally is located 5' to the initiation site for RNA polymerase II. Hundreds of promoter regions contain the "TATA" box, a sequence such as TATTA or TATAA, which is sensitive to mutations. The promoter region can be obtained by performing 5' RACE using a primer from the coding region of the gene. Alternatively, the cDNA can be used as a probe for the genomic sequence, and the region 5' to the coding region is identified by "walking up." If the gene is highly expressed or differentially expressed, the promoter from the gene can be of use in a regulatory construct for a heterologous gene.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook *et al.*, 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure herein on optional changes in amino acids to achieve altered protein structure and/or function.

As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more polynucleotides of the invention can be synthesized. Thus, the invention encompasses nucleic acid molecules ranging in length from 15 nt (corresponding to at least 15 contiguous nt of one of SEQ ID NOS:1-1079) up to a maximum length suitable for one or more biological manipulations, including replication and expression, of the nucleic acid molecule. The invention includes but is not limited to (a) nucleic acid having the size of a full gene, and comprising at least one of SEQ ID NOS:1-1079; (b) the nucleic acid of (a) also comprising at least one additional gene, operably linked to permit expression of a fusion

protein; (c) an expression vector comprising (a) or (b); (d) a plasmid comprising (a) or (b) ; and (e) a recombinant viral particle comprising (a) or (b). Once provided with the polynucleotides disclosed herein, construction or preparation of (a) - (e) are well within the skill in the art.

The sequence of a nucleic acid comprising at least 15 contiguous nt of at least any one of SEQ ID NOS:1-1079, preferably the entire sequence of at least any one of SEQ ID NOS:1-1079, is not limited and can be any sequence of A, T, G, and/or C (for DNA) and A, U, G, and/or C (for RNA) or modified bases thereof, including inosine and pseudouridine. The choice of sequence will depend on the desired function and can be dictated by coding regions desired, the intron-like regions desired, and the regulatory regions desired. Where the entire sequence of any one of SEQ ID NOS:1-1079 is within the nucleic acid, the nucleic acid obtained is referred to herein as a polynucleotide comprising the sequence of any one of SEQ ID NOS:1-1079.

Expression of Polypeptide Encoded by Full-Length cDNA or Full-Length Gene

The provided polynucleotides (*e.g.*, a polynucleotide having a sequence of one of SEQ ID NOS:1-1079), the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having sequences of SEQ ID NOS:1-1079 can also be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.*, Stemmer *et al., Gene (Amsterdam)* (1995) *164(1)*:49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) *370*:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process.

Appropriate polynucleotide constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook *et al., Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. The gene product encoded by a polynucleotide of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Vectors, host cells and methods for obtaining expression in same are well known in the art. Suitable vectors and host cells are described in USPN 5,654,173.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture.

Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. Methods for preparation of vectors comprising a desired sequence are well known in the art.

The polynucleotides set forth in SEQ ID NOS:1-1079 or their corresponding full-length polynucleotides are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence as disclosed in USPN 5,641,670.

Identification of Functional and Structural Motifs of Novel Genes Screening Against Publicly Available Databases

Translations of the nucleotide sequence of the provided polynucleotides, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the polynucleotides of the invention. Also, sequences exhibiting similarity with more than one individual sequence can exhibit activities that are characteristic of either or both individual sequences.

The full length sequences and fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence corresponding to provided polynucleotides. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences corresponding to the provided polynucleotides.

Typically, a selected polynucleotide is translated in all six frames to determine the best alignment with the individual sequences. The sequences disclosed herein in the Sequence Listing

are in a 5' to 3' orientation and translation in three frames can be sufficient (with a few specific exceptions as described in the Examples). These amino acid sequences are referred to, generally, as query sequences, which will be aligned with the individual sequences. Databases with individual sequences are described in "Computer Methods for Macromolecular Sequence Analysis" *Methods in Enzymology* (1996) *266*, Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Databases include GenBank, EMBL, and DNA Database of Japan (DDBJ).

Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST 2.0, available over the world wide web at http://www.ncbi.nlm.nih.gov/BLAST/. See also Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402. Another alignment algorithm is Fasta, available in the Genetics Computing Group (GCG) package, Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Doolittle, *supra*. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Meth. Mol. Biol.* (1997) *70*: 173-187. Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to identify sequences that are distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Amino acid sequences encoded by the provided polynucleotides can be used to search both protein and DNA databases. Incorporated herein by reference are all sequences that have been made public as of the filing date of this application by any of the DNA or protein sequence databases, including the patent databases (*e.g.*, GeneSeq). Also incorporated by reference are those sequences that have been submitted to these databases as of the filing date of the present application but not made public until after the filing date of the present application.

Results of individual and query sequence alignments can be divided into three categories: high similarity, weak similarity, and no similarity. Individual alignment results ranging from high similarity to weak similarity provide a basis for determining polypeptide activity and/or structure. Parameters for categorizing individual results include: percentage of the alignment region length where the strongest alignment is found, percent sequence identity, and p value. The percentage of the alignment region length is calculated by counting the number of residues of the individual sequence found in the region of strongest alignment, *e.g.*, contiguous region of the individual sequence that contains the greatest number of residues that are identical to the residues of the corresponding region of the aligned query sequence. This number is divided by the total residue

length of the query sequence to calculate a percentage. For example, a query sequence of 20 amino acid residues might be aligned with a 20 amino acid region of an individual sequence. The individual sequence might be identical to amino acid residues 5, 9-15, and 17-19 of the query sequence. The region of strongest alignment is thus the region stretching from residue 9-19, an 11 amino acid stretch. The percentage of the alignment region length is: 11 (length of the region of strongest alignment) divided by (query sequence length) 20 or 55%.

Percent sequence identity is calculated by counting the number of amino acid matches between the query and individual sequence and dividing total number of matches by the number of residues of the individual sequences found in the region of strongest alignment. Thus, the percent identity in the example above would be 10 matches divided by 11 amino acids, or approximately, 90.9%

P value is the probability that the alignment was produced by chance. For a single alignment, the p value can be calculated according to Karlin *et al.*, *Proc. Natl. Acad. Sci.* (1990) *87*:2264 and Karlin *et al.*, *Proc. Natl. Acad. Sci.* (1993) *90*. The p value of multiple alignments using the same query sequence can be calculated using an heuristic approach described in Altschul *et al.*, *Nat. Genet.* (1994) *6*:119. Alignment programs such as BLAST program can calculate the p value. See also Altschul *et al.*, *Nucleic Acids Res.* (1997) *25*:3389-3402.

Another factor to consider for determining identity or similarity is the location of the similarity or identity. Strong local alignment can indicate similarity even if the length of alignment is short. Sequence identity scattered throughout the length of the query sequence also can indicate a similarity between the query and profile sequences. The boundaries of the region where the sequences align can be determined according to Doolittle, *supra*; BLAST 2.0 (see, *e.g.*, Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402) or FAST programs; or by determining the area where sequence identity is highest.

High Similarity. In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. If high similarity is found, the query sequence is considered to have high similarity with a profile sequence when the p value is less than

or equal to about 10^{-2}; more usually; less than or equal to about 10^{-3}; even more usually; less than or equal to about 10^{-4}. More typically, the p value is no more than about 10^{-5}; more typically; no more than or equal to about 10^{-10}; even more typically; no more than or equal to about 10^{-15} for the query sequence to be considered high similarity.

Weak Similarity. In general, where alignment results considered to be of weak similarity, there is no minimum percent length of the alignment region nor minimum length of alignment. A better showing of weak similarity is considered when the region of alignment is, typically, at least about 15 amino acid residues in length; more typically, at least about 20; even more typically; at least about 25 amino acid residues in length. Usually, length of the alignment region can be as much as about 30 amino acid residues; more usually, as much as about 40; even more usually, as much as about 60 amino acid residues. Further, for weak similarity, the region of alignment, typically, exhibits at least about 35% of sequence identity; more typically, at least about 40%; even more typically; at least about 45% sequence identity. Usually, percent sequence identity can be as much as about 50%; more usually, as much as about 55%; even more usually, as much as about 60%.

If low similarity is found, the query sequence is considered to have weak similarity with a profile sequence when the p value is usually less than or equal to about 10^{-2}; more usually; less than or equal to about 10^{-3}; even more usually; less than or equal to about 10^{-4}. More typically, the p value is no more than about 10^{-5}; more usually; no more than or equal to about 10^{-10}; even more usually; no more than or equal to about 10^{-15} for the query sequence to be considered weak similarity.

Similarity Determined by Sequence Identity Alone. Sequence identity alone can be used to determine similarity of a query sequence to an individual sequence and can indicate the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%. Sequence identity alone as a measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, 90 residues; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in length.

Alignments with Profile and Multiple Aligned Sequences. Translations of the provided polynucleotides can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided polynucleotides can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene products (*e.g.*, polypeptides) encoded by the provided polynucleotides or corresponding cDNA or genes. For example, sequences that show an identity or similarity with a chemokine profile or MSA can exhibit chemokine activities.

Profiles can designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney *et al.*, *Nucl. Acid Res.* (1996) *24(14)*: 2730-2739. MSAs of some protein families and motifs are publicly available. For example, http://genome.wustl.edu/Pfam/ includes MSAs of 547 different families and motifs. These MSAs are described also in Sonnhammer *et al.*, *Proteins* (1997) *28*: 405-420. Other sources over the world wide web include the site at http://www.embl-heidelberg.de/argos/ali/ali.htm1; alternatively, a message can be sent to ALI@EMBL-HEIDELBERG.DE for the information. A brief description of these MSAs is reported in Pascarella *et al.*, *Prot. Eng.* (1996) *9(3)*:249-251. Techniques for building profiles from MSAs are described in Sonnhammer *et al.*, *supra*; Birney *et al.*, *supra*; and "Computer Methods for Macromolecular Sequence Analysis," *Methods in Enzymology* (1996) *266*, Doolittle, Academic Press, Inc., San Diego, California, USA.

Similarity between a query sequence and a protein family or motif can be determined by (a) comparing the query sequence against the profile and/or (b) aligning the query sequence with the members of the family or motif. Typically, a program such as Searchwise is used to compare the query sequence to the statistical representation of the multiple alignment, also known as a profile (see Birney *et al.*, *supra)*. Other techniques to compare the sequence and profile are described in Sonnhammer *et al.*, *supra* and Doolittle, *supra*.

Next, methods described by Feng *et al.*, *J. Mol. Evol.* (1987) *25*:351 and Higgins *et al.*, *CABIOS* (1989) *5*:151 can be used align the query sequence with the members of a family or motif, also known as a MSA. Sequence alignments can be generated using any of a variety of software tools. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng *et al.*, *J. Mol. Evol.* (1987) *25*:351. Another method, GAP, uses the alignment method of Needleman *et al.*, *J. Mol. Biol.* (1970) *48*:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith *et al.*, *Adv. Appl. Math.* (1981) *2*:482. In general, the following factors are used to determine if a similarity between a query sequence and a profile or

MSA exists: (1) number of conserved residues found in the query sequence, (2) percentage of conserved residues found in the query sequence, (3) number of frameshifts, and (4) spacing between conserved residues.

Some alignment programs that both translate and align sequences can make any number of frameshifts when translating the nucleotide sequence to produce the best alignment. The fewer frameshifts needed to produce an alignment, the stronger the similarity or identity between the query and profile or MSAs. For example, a weak similarity resulting from no frameshifts can be a better indication of activity or structure of a query sequence, than a strong similarity resulting from two frameshifts. Preferably, three or fewer frameshifts are found in an alignment; more preferably two or fewer frameshifts; even more preferably, one or fewer frameshifts; even more preferably, no frameshifts are found in an alignment of query and profile or MSAs.

Conserved residues are those amino acids found at a particular position in all or some of the family or motif members. Alternatively, a position is considered conserved if only a certain class of amino acids is found in a particular position in all or some of the family members. For example, the N-terminal position can contain a positively charged amino acid, such as lysine, arginine, or histidine.

Typically, a residue of a polypeptide is conserved when a class of amino acids or a single amino acid is found at a particular position in at least about 40% of all class members; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A residue is considered conserved when three unrelated amino acids are found at a particular position in the some or all of the members; more usually, two unrelated amino acids. These residues are conserved when the unrelated amino acids are found at particular positions in at least about 40% of all class member; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A query sequence has similarity to a profile or MSA when the query sequence comprises at least about 25% of the conserved residues of the profile or MSA; more usually, at least about 30%; even more usually; at least about 40%. Typically, the query sequence has a stronger similarity to a profile sequence or MSA when the query sequence comprises at least about 45% of the conserved residues of the profile or MSA; more typically, at least about 50%; even more typically; at least about 55%.

Identification of Secreted & Membrane-Bound Polypeptides

Both secreted and membrane-bound polypeptides of the present invention are of particular interest. For example, levels of secreted polypeptides can be assayed in body fluids that are convenient, such as blood, plasma, serum, and other body fluids such as urine, prostatic fluid and semen. Membrane-bound polypeptides are useful for constructing vaccine antigens or inducing an immune response. Such antigens would comprise all or part of the extracellular region of the membrane-bound polypeptides. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides.

A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods,*Proc. Natl. Acad. Sci. USA* (1981) *78*:3824-3828; Kyte & Doolittle, *J. Mol. Biol.* (1982) *157*: 105-132; and RAOAR algorithm, Degli Esposti *et al.*, *Eur. J. Biochem.* (1990) *190*: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the polynucleotides of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine

Identification of the Function of an Expression Product of a Full-Length Gene

Ribozymes, antisense constructs, and dominant negative mutants can be used to determine function of the expression product of a gene corresponding to a polynucleotide provided herein. These methods and compositions are particularly useful where the provided novel polynucleotide exhibits no significant or substantial homology to a sequence encoding a gene of known function. Antisense molecules and ribozymes can be constructed from synthetic polynucleotides. Typically, the phosphoramidite method of oligonucleotide synthesis is used. See Beaucage *et al.*, *Tet. Lett.* (1981) *22*:1859 and USPN 4,668,777. Automated devices for synthesis are available to create oligonucleotides using this chemistry. Examples of such devices include Biosearch 8600, Models 392 and 394 by Applied Biosystems, a division of Perkin-Elmer Corp., Foster City, California,

USA; and Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Synthetic RNA, phosphate analog oligonucleotides, and chemically derivatized oligonucleotides can also be produced, and can be covalently attached to other molecules. RNA oligonucleotides can be synthesized, for example, using RNA phosphoramidites. This method can be performed on an automated synthesizer, such as Applied Biosystems, Models 392 and 394, Foster City, California, USA.

Phosphorothioate oligonucleotides can also be synthesized for antisense construction. A sulfurizing reagent, such as tetraethylthiruam disulfide (TETD) in acetonitrile can be used to convert the internucleotide cyanoethyl phosphite to the phosphorothioate triester within 15 minutes at room temperature. TETD replaces the iodine reagent, while all other reagents used for standard phosphoramidite chemistry remain the same. Such a synthesis method can be automated using Models 392 and 394 by Applied Biosystems, for example.

Oligonucleotides of up to 200 nt can be synthesized, more typically, 100 nt, more typically 50 nt; even more typically 30 to 40 nt. These synthetic fragments can be annealed and ligated together to construct larger fragments. See, for example, Sambrook *et al.*, *supra*. Trans-cleaving catalytic RNAs (ribozymes) are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message must contain a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit expression of a gene of unknown function for the purpose of determining its function in an in vitro or in vivo context, by detecting the phenotypic effect. One commonly used ribozyme motif is the hammerhead, for which the substrate sequence requirements are minimal. Design of the hammerhead ribozyme, as well as therapeutic uses of ribozymes, are disclosed in Usman *et al., Current Opin. Struct. Biol.* (1996) *6*:527. Methods for production of ribozymes, including hairpin structure ribozyme fragments, methods of increasing ribozyme specificity, and the like are known in the art.

The hybridizing region of the ribozyme can be modified or can be prepared as a branched structure as described in Horn and Urdea, *Nucleic Acids Res.* (1989) *17*:6959. The basic structure of the ribozymes can also be chemically altered in ways familiar to those skilled in the art, and chemically synthesized ribozymes can be administered as synthetic oligonucleotide derivatives modified by monomeric units. In a therapeutic context, liposome mediated delivery of ribozymes improves cellular uptake, as described in Birikh *et al., Eur. J. Biochem.* (1997) *245*:1.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense polynucleotides based on a selected polynucleotide

sequence can interfere with expression of the corresponding gene. Antisense polynucleotides are typically generated within the cell by expression from antisense constructs that contain the antisense strand as the transcribed strand. Antisense polynucleotides based on the disclosed polynucleotides will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense polynucleotide. The expression products of control cells and cells treated with the antisense construct are compared to detect the protein product of the gene corresponding to the polynucleotide upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

Given the extensive background literature and clinical experience in antisense therapy, one skilled in the art can use selected polynucleotides of the invention as additional potential therapeutics. The choice of polynucleotide can be narrowed by first testing them for binding to "hot spot" regions of the genome of cancerous cells. If a polynucleotide is identified as binding to a "hot spot", testing the polynucleotide as an antisense compound in the corresponding cancer cells is warranted.

As an alternative method for identifying function of the gene corresponding to a polynucleotide disclosed herein, dominant negative mutations are readily generated for corresponding proteins that are active as homomultimers. A mutant polypeptide will interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants (see, *e.g.*, Herskowitz, *Nature* (1987) *329*:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

Polypeptides and Variants Thereof

The polypeptides of the invention include those encoded by the disclosed polynucleotides, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of SEQ ID NOS:1-1079 or a variant thereof.

In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as

the naturally occurring protein (*e.g.*, human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide of the invention, as measured by BLAST 2.0 using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, *i.e.*, the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, *i.e.* other animal or plant species, where such homologs, usually mammalian species, *e.g.* rodents, such as mice, rats; domestic animals, *e.g.*, horse, cow, dog, cat; and humans. By "homolog" is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the BLAST 2.0 algorithm, with the parameters described *supra*.

In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, *e.g.* are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/ hydrophilicity, and/or steric bulk of the amino acid substituted. Variants can be designed so as to retain or have enhanced biological activity of a particular region of the protein (*e.g.*, a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (see, e.g., Go *et al*, *Int. J. Peptide Protein Res.* (1980) *15*:211), the thermostability of the variant polypeptide (see, e.g., Querol *et al.*, *Prot. Eng.* (1996)

9:265), desired glycosylation sites (see, e.g., Olsen and Thomsen, *J. Gen. Microbiol.* (1991) *137*:579), desired disulfide bridges (see, e.g., Clarke *et al.*, *Biochemistry* (1993) *32*:4322; and Wakarchuk *et al.*, *Protein Eng.* (1994) *7*:1379), desired metal binding sites (see, e.g., Toma *et al.*, *Biochemistry* (1991) *30*:97, and Haezerbrouck *et al.*, *Protein Eng.* (1993) *6*:643), and desired substitutions with in proline loops (see, e.g., Masul *et al.*, *Appl. Env. Microbiol.* (1994) *60*:3579). Cysteine-depleted muteins can be produced as disclosed in USPN 4,959,314.

Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a polynucleotide having a sequence of any SEQ ID NOS:1-1079, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

Computer-Related Embodiments

In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (*e.g.*, as a collection of polynucleotide molecules), or in electronic form (*e.g.*, as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, *e.g.*, as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.*, cell type markers), and/or as markers of a given disease or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (*e.g.*, a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a breast ductal cell affected by cancer relative to a normal (*i.e.*, substantially disease-free) breast cell.

The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.*, electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.*, overexpressed or underexpressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a

cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of SEQ ID NOS:1-1079. By plurality is meant at least 2, usually at least 3 and can include up to all of SEQ ID NOS:1-1079. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.*, if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.* the nucleic acid sequences of any of the polynucleotides of SEQ ID NOS:1-1079, can be recorded on computer readable media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.*, searchable files, executable files, *etc*, including, but not limited to, for example, search program software, *etc.*).

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the gapped BLAST (Altschul *et al.Nucleic Acids Res.* (1997) *25*:3389-3402) and BLAZE (Brutlag *et al. Comp. Chem.* (1993) 17:203) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.* MacPattern (EMBL), BLASTN and BLASTX (NCBI). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but arc not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

As discussed above, the "library" of the invention also encompasses biochemical libraries of the polynucleotides of SEQ ID NOS:1-1079 , *e.g.*, collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.*, a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.*, an array) and the like. Of particular interest are nucleic acid arrays in which one or more of SEQ ID NOS:1-1079 is represented on the array. By array is meant a an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by SEQ ID NOS:1-1079.

Utilities

Use of Polynucleotide Probes in Mapping, and in Tissue Profiling

Polynucleotide probes, generally comprising at least 12 contiguous nt of a polynucleotide as shown in the Sequence Listing, are used for a variety of purposes, such as chromosome mapping of the polynucleotide and detection of transcription levels. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences.

Detection of Expression Levels. Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization is quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for in situ hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection

of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluors, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and USPN 5,124,246.

Alternatively, the Polymerase Chain Reaction (PCR) is another means for detecting small amounts of target nucleic acids (see, *e.g.*, Mullis *et al.*, *Meth. Enzymol.* (1987) *155*:335; USPN 4,683,195; and USPN 4,683,202). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers can be composed of sequence within or 3' and 5' to the polynucleotides of the Sequence Listing. Alternatively, if the primers are 3' and 5' to these polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art, e.g., Southern blot. mRNA or cDNA can also be detected by traditional blotting techniques (e.g., Southern blot, Northern blot, etc.) described in Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (e.g., without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

Mapping. Polynucleotides of the present invention can be used to identify a chromosome on which the corresponding gene resides. Such mapping can be useful in identifying the function of the polynucleotide-related gene by its proximity to other genes with known function. Function can also be assigned to the polynucleotide-related gene when particular syndromes or diseases map to the same chromosome. For example, use of polynucleotide probes in identification and quantification of nucleic acid sequence aberrations is described in USPN 5,783,387. An exemplary mapping method is fluorescence in situ hybridization (FISH), which facilitates comparative genomic hybridization to allow total genome assessment of changes in relative copy number of DNA sequences (see, e.g., Valdes *et al.*, *Methods in Molecular Biology* (1997) *68*:1). Polynucleotides can also be mapped to particular chromosomes using, for example, radiation hybrids or chromosome-specific hybrid panels. See Leach *et al.*, *Advances in Genetics*, (1995) *33*:63-99; Walter *et al.*, *Nature Genetics* (1994) *7*:22; Walter and Goodfellow, *Trends in Genetics* (1992) *9*:352. Panels for radiation hybrid mapping are available from Research Genetics, Inc., Huntsville, Alabama, USA. Databases for markers using various panels are available via the world wide web at http:/F/shgc-www.stanford.edu; and http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl. The statistical program RHMAP can be used to construct a map based on the data from radiation hybridization with a measure of the relative likelihood of one order versus another. RHMAP is available via the world wide web at http://www.sph.umich.edu/group/statgen/software. In addition,

commercial programs are available for identifying regions of chromosomes commonly associated with disease, such as cancer.

Tissue Typing or Profiling. Expression of specific mRNA corresponding to the provided polynucleotides can vary in different cell types and can be tissue-specific. This variation of mRNA levels in different cell types can be exploited with nucleic acid probe assays to determine tissue types. For example, PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes substantially identical or complementary to polynucleotides listed in the Sequence Listing can determine the presence or absence of the corresponding cDNA or mRNA.

Tissue typing can be used to identify the developmental organ or tissue source of a metastatic lesion by identifying the expression of a particular marker of that organ or tissue. If a polynucleotide is expressed only in a specific tissue type, and a metastatic lesion is found to express that polynucleotide, then the developmental source of the lesion has been identified. Expression of a particular polynucleotide can be assayed by detection of either the corresponding mRNA or the protein product. As would be readily apparent to any forensic scientist, the sequences disclosed herein are useful in differentiating human tissue from non-human tissue. In particular, these sequences are useful to differentiate human tissue from bird, reptile, and amphibian tissue, for example.

Use of Polymorphisms. A polynucleotide of the invention can be used in forensics, genetic analysis, mapping, and diagnostic applications where the corresponding region of a gene is polymorphic in the human population. Any means for detecting a polymorphism in a gene can be used, including, but not limited to electrophoresis of protein polymorphic variants, differential sensitivity to restriction enzyme cleavage, and hybridization to allele-specific probes.

Antibody Production

Expression products of a polynucleotide of the invention, as well as the corresponding mRNA, cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an in vitro expression system.

Methods for production of antibodies that specifically bind a selected antigen are well known in the art. Immunogens for raising antibodies can be prepared by mixing a polypeptide encoded by a polynucleotide of the invention with an adjuvant, and/or by making fusion proteins with larger immunogenic proteins. Polypeptides can also be covalently linked to other larger

immunogenic proteins, such as keyhole limpet hemocyanin. Immunogens are typically administered intradermally, subcutaneously, or intramuscularly to experimental animals such as rabbits, sheep, and mice, to generate antibodies. Monoclonal antibodies can be Monoclonal antibodies can be generated by isolating spleen cells and fusing myeloma cells to form hybridomas. Alternatively, the selected polynucleotide is administered directly, such as by intramuscular injection, and expressed in vivo. The expressed protein generates a variety of protein-specific immune responses, including production of antibodies, comparable to administration of the protein.

Preparations of polyclonal and monoclonal antibodies specific for polypeptides encoded by a selected polynucleotide are made using standard methods known in the art. The antibodies specifically bind to epitopes present in the polypeptides encoded by polynucleotides disclosed in the Sequence Listing. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. Epitopes that involve non-contiguous amino acids may require a longer polypeptide, e.g., at least 15, 25, or 50 amino acids. Antibodies that specifically bind to human polypeptides encoded by the provided polypeptides should provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in Western blots or other immunochemical assays. Preferably, antibodies that specifically polypeptides of the invention do not bind to other proteins in immunochemical assays at detectable levels and can immunoprecipitate the specific polypeptide from solution.

The invention also contemplates naturally occurring antibodies specific for a polypeptide of the invention. For example, serum antibodies to a polypeptide of the invention in a human population can be purified by methods well known in the art, e.g., by passing antiserum over a column to which the corresponding selected polypeptide or fusion protein is bound. The bound antibodies can then be eluted from the column, for example using a buffer with a high salt concentration.

In addition to the antibodies discussed above, the invention also contemplates genetically engineered antibodies, antibody derivatives (e.g., single chain antibodies, antibody fragments (e.g., Fab, etc.)), according to methods well known in the art.

Polynucleotides or Arrays for Diagnostics

Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotide sequences in a sample. This technology can be used as a diagnostic and as a tool to test for differential expression, e.g., to determine function of an encoded protein. Arrays can be created by spotting polynucleotide probes onto a substrate (*e.g.*, glass, nitrocelllose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (*e.g.*, using radioactive or fluorescent labels) and then

hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Techniques for constructing arrays and methods of using these arrays are described in EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5,593,839; USPN 5,578,832; EP 728 520; USPN 5,599,695; EP 721 016; USPN 5,556,752; WO 95/22058; and USPN 5,631,734. Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a polynucleotide between a test cell and control cell (*e.g.*, cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado *et al.*, *Sem. Radiation Oncol.* (1998) *8*:217; and Ramsay *Nature Biotechnol.* (1998) *16*:40.

Differential Expression in Diagnosis

The polynucleotides of the invention can also be used to detect differences in expression levels between two cells, *e.g.*, as a method to identify abnormal or diseased tissue in a human. For polynucleotides corresponding to profiles of protein families, the choice of tissue can be selected according to the putative biological function. In general, the expression of a gene corresponding to a specific polynucleotide is compared between a first tissue that is suspected of being diseased and a second, normal tissue of the human. The tissue suspected of being abnormal or diseased can be derived from a different tissue type of the human, but preferably it is derived from the same tissue type; for example an intestinal polyp or other abnormal growth should be compared with normal intestinal tissue. The normal tissue can be the same tissue as that of the test sample, or any normal tissue of the patient, especially those that express the polynucleotide-related gene of interest (*e.g.*, brain, thymus, testis, heart, prostate, placenta, spleen, small intestine, skeletal muscle, pancreas, and the mucosal lining of the colon). A difference between the polynucleotide-related gene, mRNA, or protein in the two tissues which are compared, for example in molecular weight, amino acid or nucleotide sequence, or relative abundance, indicates a change in the gene, or a gene which regulates it, in the tissue of the human that was suspected of being diseased. Examples of detection of differential expression and its use in diagnosis of cancer are described in USPNs 5,688,641 and 5,677,125.

A genetic predisposition to disease in a human can also be detected by comparing expression levels of an mRNA or protein corresponding to a polynucleotide of the invention in a fetal tissue with levels associated in normal fetal tissue. Fetal tissues that are used for this purpose include, but are not limited to, amniotic fluid, chorionic villi, blood, and the blastomere of an in vitro-fertilized embryo. The comparable normal polynucleotide-related gene is obtained from any

tissue. The mRNA or protein is obtained from a normal tissue of a human in which the polynucleotide-related gene is expressed. Differences such as alterations in the nucleotide sequence or size of the same product of the fetal polynucleotide-related gene or mRNA, or alterations in the molecular weight, amino acid sequence, or relative abundance of fetal protein, can indicate a germline mutation in the polynucleotide-related gene of the fetus, which indicates a genetic predisposition to disease. In general, diagnostic, prognostic, and other methods of the invention based on differential expression involve detection of a level or amount of a gene product, particularly a differentially expressed gene product, in a test sample obtained from a patient suspected of having or being susceptible to a disease (*e.g.*, breast cancer, lung cancer, colon cancer and/or metastatic forms thereof), and comparing the detected levels to those levels found in normal cells (*e.g.*, cells substantially unaffected by cancer) and/or other control cells (*e.g.*, to differentiate a cancerous cell from a cell affected by dysplasia). Furthermore, the severity of the disease can be assessed by comparing the detected levels of a differentially expressed gene product with those levels detected in samples representing the levels of differentially gene product associated with varying degrees of severity of disease. It should be noted that use of the term "diagnostic" herein is not necessarily meant to exclude "prognostic" or "prognosis," but rather is used as a matter of convenience.

The term "differentially expressed gene" is generally intended to encompass a polynucleotide that can, for example, include an open reading frame encoding a gene product (*e.g.*, a polypeptide), and/or introns of such genes and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome. In general, a difference in expression level associated with a decrease in expression level of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% or more is indicative of a differentially expressed gene of interest, *i.e.*, a gene that is underexpressed or down-regulated in the test sample relative to a control sample. Furthermore, a difference in expression level associated with an increase in expression of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% and can be at least about 1 ½-fold, usually at least about 2-fold to about 10-fold, and can be about 100-fold to about 1,000-fold increase relative to a control sample is indicative of a differentially expressed gene of interest, *i.e.*, an overexpressed or up-regulated gene.

"Differentially expressed polynucleotide" as used herein means a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product) that uniquely identifies a differentially expressed gene so that detection of the

differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotides" is also meant to encompass fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

"Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, as well as to the prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). The present invention particularly encompasses diagnosis of subjects in the context of breast cancer (*e.g.*, carcinoma in situ (*e.g.*, ductal carcinoma in situ), estrogen receptor (ER)-positive breast cancer, ER-negative breast cancer, or other forms and/or stages of breast cancer), lung cancer (*e.g.*, small cell carcinoma, non-small cell carcinoma, mesothelioma, and other forms and/or stages of lung cancer), and colon cancer (*e.g.*, adenomatous polyp, colorectal carcinoma, and other forms and/or stages of colon cancer).

"Sample" or "biological sample" as used throughout here are generally meant to refer to samples of biological fluids or tissues, particularly samples obtained from tissues, especially from cells of the type associated with the disease for which the diagnostic application is designed (*e.g.*, ductal adenocarcinoma), and the like. "Samples" is also meant to encompass derivatives and fractions of such samples (*e.g.*, cell lysates). Where the sample is solid tissue, the cells of the tissue can be dissociated or tissue sections can be analyzed.

Methods of the subject invention useful in diagnosis or prognosis typically involve comparison of the abundance of a selected differentially expressed gene product in a sample of interest with that of a control to determine any relative differences in the expression of the gene product, where the difference can be measured qualitatively and/or quantitatively. Quantitation can be accomplished, for example, by comparing the level of expression product detected in the sample with the amounts of product present in a standard curve. A comparison can be made visually; by using a technique such as densitometry, with or without computerized assistance; by preparing a representative library of cDNA clones of mRNA isolated from a test sample, sequencing the clones in the library to determine that number of cDNA clones corresponding to the same gene product, and analyzing the number of clones corresponding to that same gene product relative to the number of clones of the same gene product in a control sample; or by using an array to detect relative levels of hybridization to a selected sequence or set of sequences, and comparing the hybridization pattern to that of a control. The differences in expression are then correlated with the presence or absence

of an abnormal expression pattern. A variety of different methods for determining the nucleic acid abundance in a sample are known to those of skill in the art (see, e.g., WO 97/27317).

In general, diagnostic assays of the invention involve detection of a gene product of a the polynucleotide sequence (*e.g.*, mRNA or polypeptide) that corresponds to a sequence of SEQ ID NOS:1-1079 The patient from whom the sample is obtained can be apparently healthy, susceptible to disease (*e.g.*, as determined by family history or exposure to certain environmental factors), or can already be identified as having a condition in which altered expression of a gene product of the invention is implicated.

Diagnosis can be determined based on detected gene product expression levels of a gene product encoded by at least one, preferably at least two or more, at least 3 or more, or at least 4 or more of the polynucleotides having a sequence set forth in SEQ ID NOS:1-1079, and can involve detection of expression of genes corresponding to all of SEQ ID NOS:1-1079and/or additional sequences that can serve as additional diagnostic markers and/or reference sequences. Where the diagnostic method is designed to detect the presence or susceptibility of a patient to cancer, the assay preferably involves detection of a gene product encoded by a gene corresponding to a polynucleotide that is differentially expressed in cancer. Examples of such differentially expressed polynucleotides are described in the Examples below. Given the provided polynucleotides and information regarding their relative expression levels provided herein, assays using such polynucleotides and detection of their expression levels in diagnosis and prognosis will be readily apparent to the ordinarily skilled artisan.

Any of a variety of detectable labels can be used in connection with the various embodiments of the diagnostic methods of the invention. Suitable detectable labels include fluorochromes,(*e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (*e.g.* ^{32}P, ^{35}S, ^{3}H, *etc.*), and the like. The detectable label can involve a two stage systems (*e.g.*, biotin-avidin, hapten-anti-hapten antibody, *etc.*)

Reagents specific for the polynucleotides and polypeptides of the invention, such as antibodies and nucleotide probes, can be supplied in a kit for detecting the presence of an expression product in a biological sample. The kit can also contain buffers or labeling components, as well as instructions for using the reagents to detect and quantify expression products in the biological sample. Exemplary embodiments of the diagnostic methods of the invention are described below in more detail.

Polypeptide detection in diagnosis. In one embodiment, the test sample is assayed for the level of a differentially expressed polypeptide. Diagnosis can be accomplished using any of a number of methods to determine the absence or presence or altered amounts of the differentially expressed polypeptide in the test sample. For example, detection can utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells can be permeabilized to stain cytoplasmic molecules. In general, antibodies that specifically bind a differentially expressed polypeptide of the invention are added to a sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody can be detectably labeled for direct detection (*e.g.*, using radioisotopes, enzymes, fluorescers, chemiluminescers, and the like), or can be used in conjunction with a second stage antibody or reagent to detect binding (*e.g.*, biotin with horseradish peroxidase-conjugated avidin, a secondary antibody conjugated to a fluorescent compound, *e.g.* fluorescein, rhodamine, Texas red, *etc.*). The absence or presence of antibody binding can be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.* Any suitable alternative methods can of qualitative or quantitative detection of levels or amounts of differentially expressed polypeptide can be used, for example ELISA, western blot, immunoprecipitation, radioimmunoassay, etc.

mRNA detection. The diagnostic methods of the invention can also or alternatively involve detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotides of the invention. Any suitable qualitative or quantitative methods known in the art for detecting specific mRNAs can be used. mRNA can be detected by, for example, *in situ* hybridization in tissue sections, by reverse transcriptase-PCR, or in Northern blots containing poly A+ mRNA. One of skill in the art can readily use these methods to determine differences in the size or amount of mRNA transcripts between two samples. mRNA expression levels in a sample can also be determined by generation of a library of expressed sequence tags (ESTs) from the sample, where the EST library is representative of sequences present in the sample (Adams, et al., (1991) *Science* 252:1651). Enumeration of the relative representation of ESTs within the library can be used to approximate the relative representation of the gene transcript within the starting sample. The results of EST analysis of a test sample can then be compared to EST analysis of a reference sample to determine the relative expression levels of a selected polynucleotide, particularly a polynucleotide corresponding to one or more of the differentially expressed genes described herein. Alternatively, gene expression in a test sample can be performed using serial analysis of gene expression (SAGE) methodology (e.g., Velculescu et al., *Science* (1995) *270*:484) or differential display (DD) methodology (see, *e.g.*, U.S. 5,776,683; and U.S. 5,807,680).

Alternatively, gene expression can be analyzed using hybridization analysis. Oligonucleotides or cDNA can be used to selectively identify or capture DNA or RNA of specific sequence composition, and the amount of RNA or cDNA hybridized to a known capture sequence determined qualitatively or quantitatively, to provide information about the relative representation of a particular message within the pool of cellular messages in a sample. Hybridization analysis can be designed to allow for concurrent screening of the relative expression of hundreds to thousands of genes by using, for example, array-based technologies having high density formats, including filters, microscope slides, or microchips, or solution-based technologies that use spectroscopic analysis (*e.g.*, mass spectrometry). One exemplary use of arrays in the diagnostic methods of the invention is described below in more detail.

Use of a single gene in diagnostic applications. The diagnostic methods of the invention can focus on the expression of a single differentially expressed gene. For example, the diagnostic method can involve detecting a differentially expressed gene, or a polymorphism of such a gene (*e.g.*, a polymorphism in an coding region or control region), that is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level and/or affect activity of the encoded protein, *etc.*

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.* a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express a differentially expressed gene can be used as a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis, and a detectable label can be included in the amplification reaction (*e.g.*, using a detectably labeled primer or detectably labeled oligonucleotides) to facilitate detection. Alternatively, various methods are also known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, see e.g., Riley *et al.*, *Nucl. Acids Res.* (1990) *18*:2887; and Delahunty *et al.*, *Am. J. Hum. Genet.* (1996) *58*:1239.

The amplified or cloned sample nucleic acid can be analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy or other methods, and the sequence of bases compared to a selected sequence, *e.g.*, to a wild-type sequence. Hybridization with the polymorphic or variant sequence can also be used to determine its presence in a sample (*e.g.*, by Southern blot, dot blot, *etc.)*. The hybridization pattern of a polymorphic or variant sequence and a control sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO 95/35505, can also be used as a means of identifying polymorphic or variant sequences associated with disease. Single strand conformational

polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in a gene can be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that can affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in proteins can be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded protein can be determined by comparison with the wild-type protein.

Pattern matching in diagnosis using arrays. In another embodiment, the diagnostic and/or prognostic methods of the invention involve detection of expression of a selected set of genes in a test sample to produce a test expression pattern (TEP). The TEP is compared to a reference expression pattern (REP), which is generated by detection of expression of the selected set of genes in a reference sample (*e.g.*, a positive or negative control sample). The selected set of genes includes at least one of the genes of the invention, which genes correspond to the polynucleotide sequences of SEQ ID NOS:1-1079. Of particular interest is a selected set of genes that includes gene differentially expressed in the disease for which the test sample is to be screened.

"Reference sequences" or "reference polynucleotides" as used herein in the context of differential gene expression analysis and diagnosis/prognosis refers to a selected set of polynucleotides, which selected set includes at least one or more of the differentially expressed polynucleotides described herein. A plurality of reference sequences, preferably comprising positive and negative control sequences, can be included as reference sequences. Additional suitable reference sequences are found in GenBank, Unigene, and other nucleotide sequence databases (including, *e.g.*, expressed sequence tag (EST), partial, and full-length sequences).

"Reference array" means an array having reference sequences for use in hybridization with a sample, where the reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Usually such an array will include at least 3 different reference sequences, and can include any one or all of the provided differentially expressed sequences. Arrays of interest can further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for screening for a disease or disorder (*e.g.*, cancer, dysplasia, or other related or unrelated diseases, disorders, or conditions). The

oligonucleotide sequence on the array will usually be at least about 12 nt in length, and can be of about the length of the provided sequences, or can extend into the flanking regions to generate fragments of 100 nt to 200 nt in length or more. Reference arrays can be produced according to any suitable methods known in the art. For example, methods of producing large arrays of oligonucleotides are described in U.S. 5,134,854, and U.S. 5,445,934 using light-directed synthesis techniques. Using a computer controlled system, a heterogeneous array of monomers is converted, through simultaneous coupling at a number of reaction sites, into a heterogeneous array of polymers. Alternatively, microarrays are generated by deposition of pre-synthesized oligonucleotides onto a solid substrate, for example as described in PCT published application no. WO 95/35505.

A "reference expression pattern" or "REP" as used herein refers to the relative levels of expression of a selected set of genes, particularly of differentially expressed genes, that is associated with a selected cell type, *e.g.*, a normal cell, a cancerous cell, a cell exposed to an environmental stimulus, and the like. A "test expression pattern" or "TEP" refers to relative levels of expression of a selected set of genes, particularly of differentially expressed genes, in a test sample (*e.g.*, a cell of unknown or suspected disease state, from which mRNA is isolated).

REPs can be generated in a variety of ways according to methods well known in the art. For example, REPs can be generated by hybridizing a control sample to an array having a selected set of polynucleotides (particularly a selected set of differentially expressed polynucleotides), acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the REP with a TEP. Alternatively, all expressed sequences in a control sample can be isolated and sequenced, *e.g.*, by isolating mRNA from a control sample, converting the mRNA into cDNA, and sequencing the cDNA. The resulting sequence information roughly or precisely reflects the identity and relative number of expressed sequences in the sample. The sequence information can then be stored in a format (*e.g.*, a computer-readable format) that allows for ready comparison of the REP with a TEP. The REP can be normalized prior to or after data storage, and/or can be processed to selectively remove sequences of expressed genes that are of less interest or that might complicate analysis (*e.g.*, some or all of the sequences associated with housekeeping genes can be eliminated from REP data).

TEPs can be generated in a manner similar to REPs, *e.g.*, by hybridizing a test sample to an array having a selected set of polynucleotides, particularly a selected set of differentially expressed polynucleotides, acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the TEP with a REP. The REP and TEP to be used in a comparison can be generated simultaneously, or the TEP can be compared to previously generated and stored REPs.

In one embodiment of the invention, comparison of a TEP with a REP involves hybridizing a test sample with a reference array, where the reference array has one or more reference sequences for use in hybridization with a sample. The reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Hybridization data for the test sample is acquired, the data normalized, and the produced TEP compared with a REP generated using an array having the same or similar selected set of differentially expressed polynucleotides. Probes that correspond to sequences differentially expressed between the two samples will show decreased or increased hybridization efficiency for one of the samples relative to the other.

Methods for collection of data from hybridization of samples with a reference arrays are well known in the art. For example, the polynucleotides of the reference and test samples can be generated using a detectable fluorescent label, and hybridization of the polynucleotides in the samples detected by scanning the microarrays for the presence of the detectable label using, for example, a microscope and light source for directing light at a substrate. A photon counter detects fluorescence from the substrate, while an x-y translation stage varies the location of the substrate. A confocal detection device that can be used in the subject methods is described in USPN 5,631,734. A scanning laser microscope is described in Shalon et al., *Genome Res.* (1996) *6*:639. A scan, using the appropriate excitation line, is performed for each fluorophore used. The digital images generated from the scan are then combined for subsequent analysis. For any particular array element, the ratio of the fluorescent signal from one sample (*e.g.*, a test sample) is compared to the fluorescent signal from another sample (*e.g.*, a reference sample), and the relative signal intensity determined.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, *i.e.* data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the targets from the remaining data. The resulting data can be displayed as an image with the intensity in each region varying according to the binding affinity between targets and probes.

In general, the test sample is classified as having a gene expression profile corresponding to that associated with a disease or non-disease state by comparing the TEP generated from the test sample to one or more REPs generated from reference samples (*e.g.*, from samples associated with cancer or specific stages of cancer, dysplasia, samples affected by a disease other than cancer, normal samples, *etc.*). The criteria for a match or a substantial match between a TEP and a REP include expression of the same or substantially the same set of reference genes, as well as expression of these reference genes at substantially the same levels (*e.g.*, no significant difference between the

samples for a signal associated with a selected reference sequence after normalization of the samples, or at least no greater than about 25% to about 40% difference in signal strength for a given reference sequence. In general, a pattern match between a TEP and a REP includes a match in expression, preferably a match in qualitative or quantitative expression level, of at least one of, all or any subset of the differentially expressed genes of the invention.

Pattern matching can be performed manually, or can be performed using a computer program. Methods for preparation of substrate matrices (*e.g.*, arrays), design of oligonucleotides for use with such matrices, labeling of probes, hybridization conditions, scanning of hybridized matrices, and analysis of patterns generated, including comparison analysis, are described in, for example, U.S. 5,800,992.

Diagnosis, Prognosis and Management of Cancer

The polynucleotides of the invention and their gene products are of particular interest as genetic or biochemical markers (e.g., in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions. For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting and gene therapy. Determining expression of certain polynucleotides and comparison of a patients profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the polynucleotides of the invention are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

The polynucleotides of the invention can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. Furthermore, a polynucleotide of the invention identified as important for one type of cancer can also have implications for development or risk of development of other types of cancer, e.g., where a polynucleotide is differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide that has clinical implications for metastatic colon cancer can also have clinical implications for stomach cancer or endometrial cancer.

Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Staging assists the physician in determining a prognosis, planning treatment and evaluating the results of such treatment. Staging systems vary with the types of cancer, but generally involve the following "TNM" system: the type of tumor, indicated by T; whether the cancer has metastasized to nearby lymph nodes, indicated by N; and whether the cancer has metastasized to more distant parts of the body, indicated by M. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes it is called Stage I. If it has spread only to the closest lymph nodes, it is called Stage II. In Stage III, the cancer has generally spread to the lymph nodes in near proximity to the site of the primary lesion. Cancers that have spread to a distant part of the body, such as the liver, bone, brain or other site, are Stage IV, the most advanced stage.

The polynucleotides of the invention can facilitate fine-tuning of the staging process by identifying markers for the aggresivity of a cancer, *e.g.* the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

Grading of cancers. Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors being more aggressive than well differentiated or low-grade tumors. The following guidelines are generally used for grading tumors: 1) GX Grade cannot be assessed; 2) G1 Well differentiated; G2 Moderately well differentiated; 3) G3 Poorly differentiated; 4) G4 Undifferentiated. The polynucleotides of the invention can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressiveness of a tumor, such as metastatic potential.

Detection of lung cancer. The polynucleotides of the invention can be used to detect lung cancer in a subject. Although there are more than a dozen different kinds of lung cancer, the two main types of lung cancer are small cell and nonsmall cell, which encompass about 90% of all lung cancer cases. Small cell carcinoma (also called oat cell carcinoma) usually starts in one of the larger bronchial tubes, grows fairly rapidly, and is likely to be large by the time of diagnosis. Nonsmall cell lung cancer (NSCLC) is made up of three general subtypes of lung cancer. Epidermoid carcinoma (also called squamous cell carcinoma) usually starts in one of the larger bronchial tubes

and grows relatively slowly. The size of these tumors can range from very small to quite large. Adenocarcinoma starts growing near the outside surface of the lung and can vary in both size and growth rate. Some slowly growing adenocarcinomas are described as alveolar cell cancer. Large cell carcinoma starts near the surface of the lung, grows rapidly, and the growth is usually fairly large when diagnosed. Other less common forms of lung cancer are carcinoid, cylindroma, mucoepidermoid, and malignant mesothelioma.

The polynucleotides of the invention, e.g., polynucleotides differentially expressed in normal cells versus cancerous lung cells (e.g., tumor cells of high or low metastatic potential) or between types of cancerous lung cells (e.g., high metastatic versus low metastatic), can be used to distinguish types of lung cancer as well as identifying traits specific to a certain patient's cancer and selecting an appropriate therapy. For example, if the patient's biopsy expresses a polynucleotide that is associated with a low metastatic potential, it may justify leaving a larger portion of the patient's lung in surgery to remove the lesion. Alternatively, a smaller lesion with expression of a polynucleotide that is associated with high metastatic potential may justify a more radical removal of lung tissue and/or the surrounding lymph nodes, even if no metastasis can be identified through pathological examination.

Detection of breast cancer. The majority of breast cancers are adenocarcinomas subtypes, which can be summarized as follows: 1) ductal carcinoma in situ (DCIS), including comedocarcinoma; 2) infiltrating (or invasive) ductal carcinoma (IDC); 3) lobular carcinoma in situ (LCIS); 4) infiltrating (or invasive) lobular carcinoma (ILC); 5) inflammatory breast cancer; 6) medullary carcinoma; 7) mucinous carcinoma; 8) Paget's disease of the nipple; 9) Phyllodes tumor; and 10) tubular carcinoma;

The expression of polynucleotides of the invention can be used in the diagnosis and management of breast cancer, as well as to distinguish between types of breast cancer. Detection of breast cancer can be determined using expression levels of any of the appropriate polynucleotides of the invention, either alone or in combination. Determination of the aggressive nature and/or the metastatic potential of a breast cancer can also be determined by comparing levels of one or more polynucleotides of the invention and comparing levels of another sequence known to vary in cancerous tissue, *e.g.* ER expression. In addition, development of breast cancer can be detected by examining the ratio of expression of a differentially expressed polynucleotide to the levels of steroid hormones (*e.g.*, testosterone or estrogen) or to other hormones (*e.g.*, growth hormone, insulin). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous breast tissue, to discriminate between breast cancers with different cells of origin, to discriminate between breast cancers with different potential metastatic rates, etc.

Detection of colon cancer. The polynucleotides of the invention exhibiting the appropriate expression pattern can be used to detect colon cancer in a subject. Colorectal cancer is one of the most common neoplasms in humans and perhaps the most frequent form of hereditary neoplasia. Prevention and early detection are key factors in controlling and curing colorectal cancer. Colorectal cancer begins as polyps, which are small, benign growths of cells that form on the inner lining of the colon. Over a period of several years, some of these polyps accumulate additional mutations and become cancerous. Multiple familial colorectal cancer disorders have been identified, which are summarized as follows: 1) Familial adenomatous polyposis (FAP); 2) Gardner's syndrome; 3) Hereditary nonpolyposis colon cancer (HNPCC); and 4) Familial colorectal cancer in Ashkenazi Jews. The expression of appropriate polynucleotides of the invention can be used in the diagnosis, prognosis and management of colorectal cancer. Detection of colon cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression. Determination of the aggressive nature and/or the metastatic potential of a colon cancer can be determined by comparing levels of one or more polynucleotides of the invention and comparing total levels of another sequence known to vary in cancerous tissue, *e.g.*, expression of p53, DCC ras, lor FAP (see, e.g., Fearon ER, *et al.*, *Cell* (1990) *61(5)*:759; Hamilton SR *et al.*, *Cancer* (1993) *72*:957; Bodmer W, *et al.*, *Nat Genet.* (1994) *4(3)*:217; Fearon ER, *Ann N Y Acad Sci.* (1995) *768*:101). For example, development of colon cancer can be detected by examining the ratio of any of the polynucleotides of the invention to the levels of oncogenes (*e.g.* ras) or tumor suppressor genes (*e.g.* FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous colon tissue, to discriminate between colon cancers with different cells of origin, to discriminate between colon cancers with different potential metastatic rates, etc.

Detection of prostate cancer. The polynucleotides and their corresponding genes and gene products exhibiting the appropriate differential expression pattern can be used to detect prostate cancer in a subject. Over 95% of primary prostate cancers are adenocarcinomas. Signs and symptoms may include: frequent urination, especially at night, inability to urinate, trouble starting or holding back urination, a weak or interrupted urine flow and frequent pain or stiffness in the lower back, hips or upper thighs.

Many of the signs and symptoms of prostate cancer can be caused by a variety of other non-cancerous conditions. For example, one common cause of many of these signs and symptoms is a condition called benign prostatic hypertrophy, or BPH. In BPH, the prostate gets bigger and may block the flow or urine or interfere with sexual function. The methods and compositions of the invention can be used to distinguish between prostate cancer and such non-cancerous conditions. The methods of the invention can be used in conjunction with conventional methods of diagnosis,

e.g., digital rectal exam and/or detection of the level of prostate specific antigen (PSA), a substance produced and secreted by the prostate.

Use of Polynucleotides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full length genes can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (see, e.g., USPN 5,010,175 , and WO 91/17823). Agonists or antagonists of the polypeptides if the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, etc. The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a novel polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide of the invention, and at least one peptide agonist or antagonist of the novel binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the novel receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

Pharmaceutical Compositions and Therapeutic Uses

Pharmaceutical compositions of the invention can comprise polypeptides, antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is

determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies ora polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g., mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in *Remington's Pharmaceutical Sciences* (Mack Pub. Co., N.J. 1991).

Delivery Methods. Once formulated, the compositions of the invention can be (1) administered directly to the subject (*e.g.*, as polynucleotide or polypeptides); or (2) delivered ex vivo, to cells derived from the subject (*e.g.*, as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, e.g., subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

Methods for the ex vivo delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.*, International Publication No. WO 93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by, for example, dextran-mediated

transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Once a gene corresponding to a polynucleotide of the invention has been found to correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the disorder can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (e.g., antisense, ribozyme, etc.).

The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic compositions agents of the invention includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. Preferably, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide disclosed herein. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

Receptor-mediated targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis *et al.*, *Trends Biotechnol.* (1993) *11*:202; Chiou *et al.*, *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer* (J.A. Wolff, ed.) (1994); Wu *et al.*, *J. Biol. Chem.* (1988) *263*:621; Wu *et al.*, *J. Biol. Chem.* (1994) *269*:542; Zenke *et al.*, *Proc. Natl. Acad. Sci. (USA)* (1990) *87*:3655; Wu *et al.*, *J. Biol. Chem.* (1991) *266*:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 μg of DNA can also be used during a

gene therapy protocol. Factors such as method of action (e.g., for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect. For polynucleotide related genes encoding polypeptides or proteins with anti-inflammatory activity, suitable use, doses, and administration are described in USPN 5,654,173.

The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, *Cancer Gene Therapy* (1994) *1*:51; Kimura, *Human Gene Therapy* (1994) *5*:845; Connelly, *Human Gene Therapy* (1995) *1*:185; and Kaplitt, *Nature Genetics* (1994) *6*:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, e.g., WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; USPN 5, 219,740; WO 93/11230; WO 93/10218; USPN 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (e.g., Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, e.g., WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus as described in Curiel, *Hum. Gene Ther.* (1992) *3*:147 can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, e.g., Curiel, *Hum. Gene Ther.* (1992) *3*:147); ligand-linked DNA(see, e.g., Wu, *J. Biol. Chem.* (1989) *264*:16985); eukaryotic cell delivery vehicles cells (see, e.g., USPN 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and USPN 5,580,859. Liposomes that can act as gene delivery vehicles are described in USPN 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, *Mol. Cell Biol.* (1994) *14*:2411, and in

Woffendin, *Proc. Natl. Acad. Sci.* (1994) *91*:1581

Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al.*, *Proc. Natl. Acad. Sci. USA* (1994) *91*(24):11581. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation (see, e.g., USPN 5,206,152 and WO 92/11033). Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun (see, e.g., USPN 5,149,655); use of ionizing radiation for activating transferred gene (see, e.g., USPN 5,206,152 and WO 92/11033).

The present invention will now be illustrated by reference to the following examples which set forth particularly advantageous embodiments. However, it should be noted that these embodiments are illustrative and are not to be construed as restricting the invention in any way.

EXAMPLES

The following examples are offered primarily for purposes of illustration. It will be readily apparent to those skilled in the art that the formulations, dosages, methods of administration, and other parameters of this invention may be further modified or substituted in various ways without departing from the spirit and scope of the invention.

Example 1: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

cDNA libraries were constructed from either human colon cancer cell line Km12L4-A (Morikawa, et al., *Cancer Research* (1988) *48*:6863), KM12C (Morikawa et al. *Cancer Res.* (1988) *48*:1943-1948), or MDA-MB-231 (Brinkley et al. *Cancer Res.* (1980) *40*:3118-3129) was used to construct a cDNA library from mRNA isolated from the cells. Sequences expressed by these cell lines were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246). The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice

consistent with breast carcinoma.

The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams *et al.*, eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie *et al. Comput. Chem.* (1993) 17:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of 43 sequences. The remaining sequences were then used in a BLASTN vs. GenBank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40} were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the GenBank search), (2) weak similarity (greater than 45% identity and p value of less than 1×10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1×10^{-5}). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1×10^{-40} were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1×10^{-40} were discarded. Sequences with a p value of less than 1×10^{-65} when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1×10^{-40}, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1×10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 982 sequences listed as SEQ ID NOS:1-982 in the accompanying Sequence Listing and summarized in Table 1A (inserted prior

to claims). Each identified polynucleotide represents sequence from at least a partial mRNA transcript.

Table 1A provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the filing date of the U.S. priority application in which the sequence was first filed; 3) the attorney docket number assigned to the priority application (for internal use); 4) the SEQ ID NO assigned to the sequence in the priority application; 5) the sequence name used as an internal identifier of the sequence; and 6) the name assigned to the clone from which the sequence was isolated. Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

In order to confirm the sequences of SEQ ID NOS:1-982, the clones were retrieved from a library using a robotic retrieval system, and the inserts of the retrieved clones re-sequenced. These "validation" sequences are provided as SEQ ID NOS:983-996 in the Sequence Listing, and a summary of the "validation" sequences provided in Table 1B (inserted prior to claims). Table 1B provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the sample name assigned to the "validation" sequence obtained; and 3) the name of the clone that contains the indicated "validation" sequence. "Validation" sequences can be correlated with the original sequences they validate by referring to Table 1A. Because the "validation" sequences are often longer than the original polynucleotide sequences and thus provide additional sequence information. All validation sequences can be obtained either from the corresponding clone or from a cDNA library described herein (*e.g.*, using primers designed from the sequence provided in the sequence listing).

Example 2: Results of Public Database Search to Identify Function of Gene Products

SEQ ID NOS:1-1079 were translated in all three reading frames, and the nucleotide sequences and translated amino acid sequences used as query sequences to search for homologous sequences in either the GenBank (nucleotide sequences) or Non-Redundant Protein (amino acid sequences) databases. Query and individual sequences were aligned using the BLAST 2.0 programs, available over the world wide web at http://ww.ncbi.nlm.nih.gov/BLAST/. (see also Altschul, et al. *Nucleic Acids Res.* (1997) 25:3389-3402). The sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above in Example 1.

Tables 2A and 2B (inserted before the claims) provide the alignment summaries having a p

value of 1 x 10^{-2} or less indicating substantial homology between the sequences of the present invention and those of the indicated public databases. Table 2A provides the SEQ ID NO of the query sequence, the accession number of the GenBank database entry of the homologous sequence, and the p value of the alignment. Table 2A provides the SEQ ID NO of the query sequence, the accession number of the Non-Redundant Protein database entry of the homologous sequence, and the p value of the alignment. The alignments provided in Tables 2A and 2B are the best available alignment to a DNA or amino acid sequence at a time just prior to filing of the present specification. The activity of the polypeptide encoded by the SEQ ID NOS listed in Tables 2A and 2B can be extrapolated to be substantially the same or substantially similar to the activity of the reported nearest neighbor or closely related sequence. The accession number of the nearest neighbor is reported, providing a publicly available reference to the activities and functions exhibited by the nearest neighbor. The public information regarding the activities and functions of each of the nearest neighbor sequences is incorporated by reference in this application. Also incorporated by reference is all publicly available information regarding the sequence, as well as the putative and actual activities and functions of the nearest neighbor sequences listed in Table 2 and their related sequences. The search program and database used for the alignment, as well as the calculation of the p value are also indicated.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of the corresponding polynucleotide. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of the corresponding polynucleotides.

Example 3: Identification of Contiguous Sequences Having a Polynucleotide of the Invention

The novel polynucleotides were used to screen publicly available and proprietary databases to determine if any of the polynucleotides of SEQ ID NOS:1-982 would facilitate identification of a contiguous sequence, *e.g.*, the polynucleotides would provide sequence that would result in 5' extension of another DNA sequence, resulting in production of a longer contiguous sequence composed of the provided polynucleotide and the other DNA sequence(s). Contiging was performed using the Gelmerge application (default settings) of GCG from the Univ. of Wisconsin.

Using these parameters, 83 contiged sequences were generated. These contiged sequences are provided as SEQ ID NOS:997-1079 (see Table 1C). Table 1C provides the SEQ ID NO of the contig sequence, the name of the sequence used to create the contig, and the accession number of the publicly available tentative human consensus (THC) sequence used with the sequence of the corresponding sequence name to provide the contig. The sequence name of Table 1C can be

correlated with the SEQ ID NO: of the polynucleotide used to generate the contig by referring to Tables 1A and 1B.

The contiged sequences (SEQ ID NOS:997-1079)represent longer sequences that encompass another of the polynucleotide sequence of the invention. The contiged sequences were then translated in all three reading frames to determine the best alignment with individual sequences using the BLAST programs as described above. The sequences were masked using the XBLAST program for masking low complexity as described above in Example 1. As described in more detail below, several of the contiged sequences were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (see Example 4 and Table 3 below). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Example 4: Members of Protein Families

SEQ ID NOS:1-1079 were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein family (and thus represent nmembers of these protein families) and/or comprising a known functional domain. Table 3 (inserted before claims) provides the SEQ ID NO: of the query sequence, a brief description of the profile hit, the position of the query sequence within the individual sequence (indicated as "start" and "stop"), and the orientation (Direction, "Dir") of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits where the query sequence contains overlapping profile regions, and/or where the sequence contains two different functional domains. Each of the profile hits of Table 3 are described in more detail below. The acronyms for the profiles (provided in parentheses) are those used to identify the profile in the Pfam and Prosite databases. The Pfam database can be accessed through any of the following URLS: http://pfam.wustl.edu/index.html; http://www.sanger.ac.uk/ Software/Pfam/ and http://www.cgr.ki.se/Pfam/. The Prosite database can be accessed at http://www.expasy.ch/prosite/ The public information available on the Pfam and Prosite databases regarding the various profiles, including but not limited to the activities, function, and consensus sequences of various proteins families and protein domains, is incorporated herein by reference.

14-3-3 Family (14_3_3; Pfam Pfam Accession No. PF00244). SEQ ID NO:1053 corresponds to a sequence encoding a 14-3-3 protein family member. The 14-3-3 protein family includes a group of closely related acidic homodimeric proteins of about 30 kD first identified as very abundant in mammalian brain tissues and located preferentially in neurons (Aitken et al. *Trends Biochem. Sci.* (1995) *20*:95-97; Morrison *Science* (1994) *266*:56-57; and Xiao et al. *Nature* (1995) *376*:188-191). The 14-3-3 proteins have multiple biological activities, including a key role in signal transduction pathways and the cell cycle. 14-3-3 proteins interact with kinases (*e.g.*, PKC or Raf-1), and can also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases. The 14-3-3 protein sequences are extremely well conserved, and include two highly conserved regions: the first is a peptide of 11 residues located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section. The consensus patterns are as follows: 1) R-N-L-[LIV]-S-[VG]-[GA]-Y-[KN]-N-[IVA]; 2) Y-K-[DE]-S-T-L-I-[IM]-Q-L-[LF]-[RHC]-D-N-[LF]-T-[LS]-W-[TAN]-[SAD].

Ank Repeats (ANK; Pfam Accession No. PF0023). SEQ ID NO:311, represents a polynucleotide encoding an Ank repeat-containing protein. The ankyrin motif is a 33 amino acid sequence named after the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breeden *et al.Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno *et al., Development* (1997) *124(21)*:4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (Biochem J. 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) *17(4)*:363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr *et al., Current Op. Cell Biol.* (1992) *4*:496; Bennet *et al., J. Biol. Chem.* (1980) *255*:6424).

ATPases Associated with Various Cellular Activities (ATPases; Pfam Accession No. PF0004). SEQ ID NOS:1035, 1058, and 1072 corresond to a sequence that encodes a member of a family of ATPases Associated with diverse cellular Activities (AAA). The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids containing an ATP-binding site (Froehlich *et al.,J. Cell Biol.* (1991) *114*:443; Erdmann *et al. Cell* (1991) *64*:499; Peters *et al., EMBO J.* (1990) *9*:1757; Kunau *et al., Biochimie* (1993) *75*:209-224; Confalonieri *et al., BioEssays* (1995) *17*:639; http://yeamob.pci. chemie.uni-tuebingen.de/AAA/Description.html). The AAA domain, which can be present in one or two copies, acts as an ATP-dependent protein clamp (Confalonieri *et al.* (1995) *BioEssaysl 7*:639) and contains a highly conserved region located in the central part of the domain. The consensus pattern is: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]- D-x-A-[LIFA]-x-R.

Basic Region Plus Leucine Zipper Transcription Factors (BZIP; Pfam Accession No. PF00170). SEQ ID NO:918 represents a polynucleotide encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. The consensus pattern for this protein family is: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

EF Hand (Efhand; Pfam Accession No. PF00036). SEQ ID NO:242 corresponds to a polynucleotide encoding a member of the EF-hand protein family, a calcium binding domain shared by many calcium-binding proteins belonging to the same evolutionary family (Kawasaki *et al.*, *Protein. Prof.* (1995) *2*:305-490). The domain is a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain, with a calcium ion coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand). The consensus pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW].

Ets Domain (Ets_Nterm; Pfam Accession No. PF110178). SEQ ID NO:547, and thus the sequence it validates, represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., *Eur. J. Biochem.* (1993) *211*:718). The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence and comprises an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

(FKH; Pfam Accession No.PF00250). SEQ ID NO:925 corresponds to a gene encoding a polypeptide comprising a forkhead domain. The forkhead domain (also known as a "winged helix") is present in a family of eukaryotic transcription factors, and is a conserved domain of about 100 amino acid residues that is involved in DNA-binding (Weigel *et al.Cell* (1990) *63*:455-456;

Clark *et al. Nature* (1993) *364*:412-420). Mammalian genes that comprise a forkhead domain include those encoding: 1) transcriptional activators (*e.g.*, HNF-3-alpha, -beta, and -gamma proteins, which interact with the cis-acting regulatory regions of a number of liver genes); 2) interleukin-enhancer binding factor (ILF), which binds to purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter and is involved in both positive and negative regulation of important viral and cellular promoter elements; 3) transcription factor BF-1, which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon; 4) human HTLF, which binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR); 5) transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4; 6) human AFX1 which is involved in a chromosomal translocation that causes acute leukemia; and 7) human FKHR which is involved in a chromosomal translocation that causes rhabdomyosarcoma. The fork domain is highly conserved, and is detected by two consensus patterns: the first corresponding to the N-terminal section of the domain; the second corresponding to a heptapeptide located in the central section of the domain. The consensus patterns are as follows: 1) [KR]-P-[PTQ]-[FYLVQH]-S-[FY]-x(2)-[LIVM]-x(3,4)-[AC]- [LIM]; and 2) W-[QKR]-[NS]-S-[LIV]-R-H.

Helicases conserved C-terminal domain (helicase_C; Pfam Accession No. PF00271). SEQ ID NOS:227 and 1058 represent polynucleotides encoding novel members of the DEAD/H helicase family. The DEAD box family comprises a number of eukaryotic and prokaryotic proteins involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman, *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578; http://www.expasy.ch/www/linder/ HELICASES_TEXT.html). One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins (Wassarman D.A., et al., *Nature* (1991) *349*:463; Harosh I., et al., *Nucleic Acids Res.* (1991) *19*:6331; Koonin E.V., et al., *J. Gen. Virol.* (1992) *73*:989; http://www.expasy.ch/www/linder/ HELICASES_TEXT.html). The following signature patterns are used to identify member for both subfamilies: 1) [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]; and 2) [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

Kazal serine protease inhibitors family signature (Kazal; Pfam Accession No. PF00050). SEQ ID NO:97 corresponds to a polynucleotide of a gene encoding a serine protease inhibitor of the

Kazal inhibitor family (Laskowski *et al.Annu. Rev. Biochem.* (1980) *49*:593-626). The basic structure of Kazal serine protease inhibitors such a type of inhibitor is described at Pfam Accession No. PF00050. Exemplary proteins known to belong to this family include: pancreatic secretory trypsin inhibitor (PSTI), whose physiological function is to prevent the trypsin-catalyzed premature activation of zymogens within the pancreas; mammalian seminal acrosin inhibitors; canidae and felidae submandibular gland double-headed protease inhibitors, which contain two Kazal-type domains, the first one inhibits trypsin and the second one elastase; a mouse prostatic secretory glycoprotein, induced by androgens, and which exhibits anti-trypsin activity; avian ovomucoids; chicken ovoinhibitor; and the leech trypsin inhibitor Bdellin B-3. The consensus pattern is as follows: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C, where the four C's are involved in disulfide bonds.

MAP kinase kinase (mkk). SEQ ID NOS:635 and 992 represent members of the MAP kinase kinase (mkk) family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKKs are essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida *et al.*, *Trends Biochem Sci* (1993) *18*:128-31; Ruderman *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran *et al.*, *Oncogene* (1998) *17*:1447-55; Kiefer *et al.*, *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Neurotransmitter-Gated Ion-Channel (neur_chan; Pfam Accession No. PF00065). SEQ ID NO:1078 corresponds to a sequence encoding a neurotransmitter-gated ion channel. Neurotransmitter-gated ion-channels, which provide the molecular basis for rapid signal transmission at chemical synapses, are post-synaptic oligomeric transmembrane complexes that transiently form a ionic channel upon the binding of a specific neurotransmitter. Five types of neurotransmitter-gated receptors are known: 1) nicotinic acetylcholine receptor (AchR); 2) glycine receptor; 3) gamma-aminobutyric-acid (GABA) receptor; 4) serotonin 5HT3 receptor; and 5) glutamate receptor. All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related, and are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions that form the ionic channel, followed by an intracellular region of variable length. A fourth hydrophobic region is found at the C-terminal of the sequence. The consensus pattern is: C-x-[LIVMFQ]-x-[LIVMF]-x(2)-[FY]-P-x-D-x(3)-C, where the two C's are linked by a disulfide bond.

PDZ Domain (PDZ; Pfam Accession No. PF00595.) SEQ ID NOS:523 and 980 correspond to a gene comprising a PDZ domain (also known as DHR or GLGF domain). PDZ domains comprise 80-100 residue repeats, several of which interact with the C-terminal tetrapeptide motifs X-Ser/Thr-X-Val-COO- of ion channels and/or receptors, and are found in mammalian proteins as well as in bacteria, yeast, and plants (Pontig *et al.Protein Sci* (1997) *6*(2):464-8). Proteins comprising one or more PDZ domains are found in diverse membrane-associated proteins, including members of the MAGUK family of guanylate kinase homologues, several protein phosphatases and kinases, neuronal nitric oxide synthase, and several dystrophin-associated proteins, collectively known as syntrophins (Ponting *et al. Bioessays* (1997) *19*(6):469-79). Many PDZ domain-containing proteins are localised to highly specialised submembranous sites, suggesting their participation in cellular junction formation, receptor or channel clustering, and intracellular signalling events. For example, PDZ domains of several MAGUKs interact with the C-terminal polypeptides of a subset of NMDA receptor subunits and/or with Shaker-type K+ channels. Other PDZ domains have been shown to bind similar ligands of other transmembrane receptors. In cell junction-associated proteins,the PDZ mediates the clustering of membrane ion channels by binding to their C-terminus. The X-ray crystallographic structure of some proteins comrpising PDZ domains have been solved (see, *e.g.*, Doyle *et al.Cell* (1996) *85*(7):1067-76).

Protein phosphatase 2A regulatory subunit PR55 signatures (PR55; Pfam Accession No. PF01240). SEQ ID NO:1028 corresponds to a gene encoding a protine phosphatase 2A reguatory subunit. Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric enzyme that consists of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit (PR65), also called subunit A; this complex then associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme (Mayer *et al. Trends Cell Biol.* (1994) *4*:287-291). One of the forms of the variable subunit is a 55 Kd protein (PR55) which is highly conserved in mammals (where three isoforms are known to exist). This subunit may perform a substrate recognition function or be responsible for targeting the enzyme complex to the appropriate subcellular compartment. Two perfectly conserved sequences of 15 residues, one located the N-terminal region, the other in the center of the protein, serve as the basis for the consensus patterns: 1) E-F-D-Y-L-K-S-L-E-I-E-E-K-I-N; 2) N-[AG]-H-[TA]-Y-H-I-N-S-I-S-[LIVM]-N-S-D

Protein Kinase (protkinase; Pfam Accession No. PF00069). SEQ ID NOS:635, 992, and 1078 represent polynucleotides encoding protein kinases, which catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks, *et al.*, *FASEB J.* (1995) *9*:576; Hunter, *Meth. Enzymol.* (1991) *200*:3; Hanks, *et al.*, *Meth. Enzymol.*

(1991) *200*:38; Hanks, *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks *et al.*, *Science* (1988) *241*:42) belong to a very extensive family of proteins that share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The first region, located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, located in the central part of the catalytic domain, contains a conserved an aspartic acid residue that is important for the catalytic activity of the enzyme (Knighton, *et al.*, *Science* (1991) *253*:407).

The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks, *et al.*, *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain. The consensus patterns are as follows: 1) [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K, where K binds ATP; 2) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3), where D is an active site residue; and 3) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC], where D is an active site residue.

Ras family proteins (ras; Pfam Accession No. PF00071). SEQ ID NO:527 represents polynucleotides encoding the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, Biochemistry 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg2+ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, Biochemistry 30:4637-4648. A major consensus pattern of ras proteins is: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

Src homology domain 3 (SH3; Pfam Accession No. PF00018). SEQ IDNO:450 corresponds to a gene comprising a Src homology domain. The Src homology 3 (SH3) domain is a small protein domain of about 60 amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g. Src, Abl, Lck) (Mayer *et al. Nature* (1988) *332*:272-275). Since then, it has been found in a great variety of other intracellular or membrane-associated proteins (Musacchio *et alFEBS Lett.* (1992) *307*:55-61; Pawson *et al. Curr. Biol.* (1993) *3*:434-442; Mayer *et al. Trends Cell Biol.* (1993) *3*:8-13; Pawson *Nature* (1995) *373*:573-580). The SH3 domain has a characteristic fold which consists of five or six

beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan *et al. Curr. Opin. Struct. Biol.* (1993) *3*:828-837). The SH3 domain is thought to mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton *et al. Curr. Biol.* (1994) *4*:615-617). In general SH3 domains are found as single copies in a given protein, but there a significant number of proteins comprise two SH3 domains and a few comprise 3 or 4 copies. The profile to detect SH3 domains is based on a structural alignment consisting of 5 gap-free blocks and 4 linker regions totaling 62 match positions.

Trypsin (trypsin; Pfam Accession No. PF00089). SEQ ID NOS:635, 995, and 984 correspond to novel serine proteases of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved (Brenner *Nature* (1988) *334*:528). The consensus patterns for the trypsin protein family are: 1) [LIVM]-[ST]-A-[STAG]-H-C, where H is the active site residue; and 2) [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH], where S is the active site residue. All sequences known to belong to this family are detected by the above consensus sequences, except for 18 different proteases which have lost the first conserved glycine. If a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100%.

WD Domain, G-Beta Repeats (WD domain; Pfam Accession No. PF00400) SEQ ID NOS:505, 721, and 1018 represent a members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the beta and gamma subunits are required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta has eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). The consensus pattern for the WD domain/G-Beta repeat family is: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

WW/rsp5/WWP domain signature and profile (WW domain; Pfam Accession No. PF00397). SEQ ID NO:606 corresponds to a gene encoding a protein comprising a WW domain. The WW domain (Bork *et al. Trends Biochem. Sci.* (1994) *19*:531-533; Andre *et al. Biochem. Biophys. Res. Commun.* (1994) *205*:1201-1205; Hofmann *et al. FEBS Lett.* (1995) *358*:153-157;

Sudol *et al. FEBS Lett.* (1995) *369*:67-71; http://www.bork.embl-heidelberg.de/Modules/ww-gif.html) (also known as rsp5 or WWP) was discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown (Chen *et al. Proc. Natl. Acad. Sci. U.S.A.* (1995) *92*:7819-7823) to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. The WW domain conatins beta-strands grouped around four conserved aromatic positions, generally tryptophan. The name WW or WWP derives from the presence of two tryptophane as well as a conserved proline. The WW domain is frequently associated with other domains typical for proteins in signal transduction processes. The consensus pattern for WW domains is: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

Zinc Finger, C2H2 Type (Zincfing_C2H2; Pfam Accession No. PF00096). Several sequences corresponded to polynucleotides encoding members of the C2H2 type zinc finger protein family, which contain zinc finger domains that facilitate nucleic acid binding (Klug *et al.*, *Trends Biochem. Sci.* (1987) *12*:464; Evans *et al.*, *Cell* (1988) *52*:1; Payre *et al.*, *FEBS Lett.* (1988) *234*:245; Miller *et al.*, *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) *85*:99). In addition to the conserved zinc ligand residues, a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld *et al.*, *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position, which is generally an aromatic or aliphatic residue, is located four residues after the second cysteine. The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.

Zinc finger, C3HC4 type (RING finger), signature (Zincfing_C3H4; Pfam Accession No. PF00097). SEQ ID NOS:805 and 1078 represent polynucleotides encoding a polypeptide having a C3HC4 type zinc finger signature. A number of eukaryotic and viral proteins contain this signature, which is primarily a conserved cysteine-rich domain of 40 to 60 residues (Borden K.L.B., et al., *Curr. Opin. Struct. Biol.* (1996) *6*:395) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D structure of the zinc ligation system is uniqueto the RING domain and is refered to as the "cross-brace" motif. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The signature pattern for the C3HC4 finger is based on the central region of the domain: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

Zinc finger, CCHC type (Zincfing_CCHC; Pfam Accession No. PF00098). SEQ ID NOS:693,973, and 1078 correspond to genes encoding a member of the family of CCHC zinc fingers. Because the prototype CCHC type zinc finger structure is from an HIV protein, this domain

is also referred to as a retrovrial-type zinc finger domain. The family also contains proteins involved in eukaryotic gene regulation, such as C. elegans GLH-1. The structure is an 18-residue zinc finger; no examples of indels in the alignment. The motif that defines a CCHC type zinc finger domain is: C-X2-C-X4-H-X4-C (Summers J Cell Biochem 1991 Jan;45(1):41-8). The domain is found in, for example, HIV-1 nucleocapsid protein, Moloney murine leukemia virus nucleocapsid protine NCp10 (De Rocquigny *et al.Nucleic Acids Res.* (1993) *21*:823-9), and myelin transcription factor 1 (Myt1) (Kim *et al.J. Neurosci. Res.* (1997) *50*:272-90).

Example 5: Differential Expression of Polynucleotides of the Invention: Description of Libraries and Detection of Differential Expression

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 4 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the "nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 4. Description of cDNA Libraries

Library (lib #)	Description	**Number of Clones in Library**
1	Human Colon Cell Line Km12 L4: High Metastatic Potential (derived from Km12C)	308731
2	Human Colon Cell Line Km12C: Low Metastatic Potential	284771
3	Human Breast Cancer Cell Line MDA-MB-231: High Metastatic Potential; micro-mets in lung	326937
4	Human Breast Cancer Cell Line MCF7: Non Metastatic	318979
8	Human Lung Cancer Cell Line MV-522: High Metastatic Potential	223620
9	Human Lung Cancer Cell Line UCP-3: Low Metastatic Potential	312503
12	Human microvascular endothelial cells (HMVEC) - UNTREATED (PCR (OligodT) cDNA library)	41938
13	Human microvascular endothelial cells (HMVEC) – bFGF TREATED (PCR (OligodT) cDNA library)	42100
14	Human microvascular endothelial cells (HMVEC) – VEGF TREATED (PCR (OligodT) cDNA library)	42825
15	Normal Colon - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	282722
16	Colon Tumor - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	298831
17	Liver Metastasis from Colon Tumor of UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	303467
18	Normal Colon - UC#3 Patient (MICRODISSECTED PCR	36216

Library (lib #)	Description	Number of Clones in Library
	(OligodT) cDNA library)	
19	Colon Tumor - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	30956
21	GRRpz Cells derived from normal prostate epithelium	164801
22	WOca Cells derived from Gleason Grade 4 prostate cancer epithelium	162088
23	Normal Lung Epithelium of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	306198
24	Primary tumor, Large Cell Carcinoma of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	309349

The KM12L4, KM12C, and MDA-MB-231 cell lines are described in Example 1 above. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3); Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMVEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMVEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz and WOca cell lines were provided by Dr. Donna M. Peehl, Department of Medicine, Stanford University School of Medicine. GRRpz was derived from normal prostate epithelium. The WOca cell line is a Gleason Grade 4 cell line.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their

hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1st), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2nd). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar,

Biostatistical Analysis, Prentice Hall, Inc., USA, "Differences between Proportions," pp 296-298 (1974).

Examples 6-11: Differential Expression of Polynucleotides of the Invention

A number of polynucleotide sequences have been identified that are differentially expressed between, for example, cells derived from high metastatic potential cancer tissue and low metastatic cancer cells, and between cells derived from metastatic cancer tissue and normal tissue. Evaluation of the levels of expression of the genes corresponding to these sequences can be valuable in diagnosis, prognosis, and/or treatment (*e.g.*, to facilitate rationale design of therapy, monitoring during and after therapy, *etc.*). Moreover, the genes corresponding to differentially expressed sequences described herein can be therapeutic targets due to their involvement in regulation (*e.g.*, inhibition or promotion) of development of, for example, the metastatic phenotype. For example, sequences that correspond to genes that are increased in expression in high metastatic potential cells relative to normal or non-metastatic tumor cells may encode genes or regulatory sequences involved in processes such as angiogenesis, differentiation, cell replication, and metastasis.

Detection of the relative expression levels of differentially expressed polynucleotides described herein can provide valuable information to guide the clinician in the choice of therapy. For example, a patient sample exhibiting an expression level of one or more of these polynucleotides that corresponds to a gene that is increased in expression in metastatic or high metastatic potential cells may warrant more aggressive treatment for the patient. In contrast, detection of expression levels of a polynucleotide sequence that corresponds to expression levels associated with that of low metastatic potential cells may warrant a more positive prognosis than the gross pathology would suggest.

A number of polynucleotide sequences of the present invention are differentially expressed between human microvascular endothelial cells (HMVEC) that have been treated with growth factors relative to untreated HMVEC. Sequences that are differentially expressed between growth factor-treated HMVEC and untreated HMVEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other development and oncogenic processes. For example, sequences that are more highly expressed in HMVEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMVEC can serve as drug targets for chemotherapeutics, *e.g.*, decreasing expression of such up-regulated genes or inhibitin gthe activity of the encoded gene product would serve to inhibit tumor cell angiogenesis. Detection of expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. A patient sample

displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The differential expression of the polynucleotides described herein can thus be used as, for example, diagnostic markers, prognostic markers, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers. The following examples provide relative expression levels of polynucleotides from specified cell lines and patient tissue samples.

Example 6: High Metastatic Potential Breast Cancer Versus Low Metastatic Breast Cancer Cells

The tables bellow summarize the data for polynucleotides that represent genes differentially expressed between high metastatic potential and low metastatic potential breast cancer cells.

Table 5. High metastatic potential breast (lib3) > low metastatic potential breast cancer cells (lib4)

SEQ ID NO:	Lib 3 Clones	Lib4 Clones	Lib3/Lib4
781	13	0	12.68
778	9	0	8.78
756	8	0	7.81
779	7	0	6.83
691	7	0	6.83
686	7	0	6.83
916	6	0	5.85

Table 6. Low metastatic potential breast (lib4) > high metastatic potential breast cancer cells (lib3)

Table 6			
SEQ ID NO:	**Lib 3 Clones**	**Lib4 Clones**	**Lib4/Lib3**
558	0	340	348.48
656	0	64	65.6
661	0	57	58.42
647	0	43	44.07
547	0	41	42.02
648	0	40	41
592	4	115	29.47
654	0	28	28.7
646	0	21	21.52
636	3	61	20.84
533	1	17	17.42
549	0	17	17.42
650	3	50	17.08
589	0	16	16.4
110	0	16	16.4
657	0	16	16.4
624	0	16	16.4
637	0	13	13.32

Table 6			
SEQ ID NO:	**Lib 3 Clones**	**Lib4 Clones**	**Lib4/Lib3**
536	0	12	12.3
653	1	11	11.27
562	1	11	11.27
587	1	11	11.27
609	1	11	11.27
590	0	10	10.25
641	0	10	10.25
532	1	10	10.25
623	0	9	9.22
591	0	8	8.2
521	0	8	8.2
214	0	7	7.17
607	0	7	7.17
554	0	7	7.17
555	0	7	7.17
582	0	7	7.17
584	0	7	7.17
599	0	7	7.17
561	0	6	6.15
572	0	6	6.15
359	0	6	6.15
635	0	6	6.15
113	0	6	6.15
603	0	6	6.15

Example 7: High Metastatic Potential Lung Cancer Versus Low Metastatic Lung Cancer Cells

The following summarizes polynucleotides that represent genes differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 7. High metastatic potential lung (lib8) > low metastatic potential lung cancer cells (lib9)

SEQ ID NO:	**Lib 8 Clones**	**Lib 9 Clones**	**Lib8/Lib9**
571	35	1	48.91
969	8	0	11.18
350	5	0	6.99

Example 8: High Metastatic Potential Colon Cancer Versus Low Metastatic Colon Cancer Cells

Table 8 summarizes polynucleotides that represent genes differentially expressed between high metastatic potential and low metastatic potential colon cancer cells:

Table 8. Low metastatic potential colon (lib2) > high metastatic potential colon cancer cells (lib1)

SEQ ID NO:	**Lib1 Clones**	**Lib2 Clones**	**Lib2/Lib1**
57	0	8	8.67

103	0	6	6.5
189	0	6	6.5

Example 9: High Tumor Potential Colon Tissue Vs. Metastasized Colon Cancer Tissue

The following table summarizes polynucleotides that represent genes differentially expressed between high tumor potential colon cancer cels and cells derived from high metastatic potential colon cancer cells of a patient.

Table 9. High tumor potential colon tissue (lib16) vs. high metastatic colon tissue (lib17)

SEQ ID NO:	Lib 16 Clones	Lib 17 Clones	Lib17/Lib16
100	0	7	6.89
370	3	12	3.94

Example 10: Differential Expression Across Multiple Libraries

A number of polynucleotide sequences have been identified that represent genes that are differentially expressed across multiple libraries. Expression of these sequences in a tissue or any origin can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The differential expression data for these sequences is provided inTable 10 below.

Table 10. Genes Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
34	Low Met Colon (lib2) > High Met Colon (lib1)	8.67
34	High Met Breast (lib3) > Low Met Breast (Lib4)	5.85
209	Low Met Lung (lib9) > High Met Lung (lib8)	17.44
209	Colon Tumor Tissue (lib16) > Normal Colon Tissue (lib15)	3.42
209	Colon Tumor Tissue (lib19) > Normal Colon Tissue (lib18)	66.5
209	High Met Colon Tissue (lib20) > Normal Colon Tissue (lib18)	14.04
209	Colon Tumor Tissue (lib19) > High Met Colon Tissue (lib20)	4.74
316	High Met Colon (lib1) > Low Met Colon (lib2)	5.76
316	Low Met Breast (lib4) > High Met Breast (Lib3)	17.28
645	Low Met Breast (lib4) > High Met Breast (Lib3)	6.15

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
645	High Met Lung (lib8) > Low Met Lung (lib9)	19.56
854	High Met Breast (lib3) > Low Met Breast (Lib4)	9.76
854	HMVEC-bFGF (lib13) > HMVEC (lib12)	4.98
854	Lung Tumor Tissue (lib24) > Normal Lung Tissue (lib23)	5.94

Key for Table 10: High Met = high metastatic potential; Low Met = low metastatic potential; met = metastasized; tumor = non-metastasized tumor; HMVEC = human microvascular endothelial cell; bFGF = bFGF treated.

Detection of expression of genes that correspond to the above polynucleotides may be of particular interest in diagnosis, prognosis, risk assesment, and monitoring of treatment. Furthermore, differential expression of a specific gene across multiple libraries can also be indicative of a gene whose expression is associated with, for example, suppression of the metastatic phenotype or with development of the cell toward a metastatic phenotype. For example, SEQ ID NO:209 corresponds to a gene that is expressed at relatively higher levels in colon tumor tissue than in high metastatic potential colon tumor tissue, and at relatively higher levels in high metastatic potential colon tumor tissue than in normal colon tissue. Thus a relatively increased level of expression of the gene corresponding to SEQ ID NO:209 may be used as marker of a pre-metastatic colon cells either alone or in combination with other markers.

Some polynucleotides exhibited opposite differential expression trends in libraries of different origin (see, *e.g.*., SEQ ID NO:316). These data suggest that the differential expressio patterns of some gene associated with development of metastases indicate a unique role for those genes specific for the tissue of origin.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration

and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information. The following materials were deposited with the American Type Culture Collection (CMCC = Chiron Master Culture Collection).

Table 11. Cell Lines Deposited with ATCC

Cell Line	Deposit Date	ATCC Accession No.	CMCC Accession No.
KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

In addition, pools of selected clones, as well as libraries containing specific clones, were assigned an "ES" number (internal reference) and deposited with the ATCC. Table 21 below provides the ATCC Accession Nos. of the ES deposits, all of which were deposited on or before May 13, 1999. The names of the clones contained within each of these deposits are provided in the tables numbered 22 and greater (inserted before the claims).

Table 12: Pools of Clones and Libraries Deposited with ATCC on or before September 23, 1999

Library No.	CMCC No.	ATCC Deposit No.	Library No.	CMCC No.	ATCC Deposit No.
ES55	5058		ES65	5068	
ES56	5059		ES66	5069	
ES57	5060		ES67	5070	
ES58	5061		ES68	5071	
ES59	5062		ES69	5072	
ES60	5063		ES70	5073	
ES61	5064		ES71	5074	
ES62	5065		ES72	5075	
ES63	5066		ES73	5076	
ES64	5067		ES74	5077	

The deposits described herein are provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Retrieval of Individual Clones from Deposit of Pooled Clones. Where the ATCC deposit is composed of a pool of cDNA clones or a library of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones in the pool or library were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
1	9/28/98	1492.001	1	RTA00000617F.o.18.2	M00005513A:H01
2	9/28/98	1492.001	2	RTA00001075F.h.12.1	M00005434A:F11
3	9/28/98	1492.001	3	RTA00001076F.m.09.1	M00006946B:C08
4	9/28/98	1492.001	4	RTA00001075F.o.08.1	M00005628D:A10
5	9/28/98	1492.001	5	RTA00001064F.f.14.1	M00005465A:A07
6	9/28/98	1492.001	6	RTA00001075F.n.19.1	M00005614A:B07
7	9/28/98	1492.001	7	RTA00001075F.i.24.1	M00005453B:B06
8	9/28/98	1492.001	8	RTA00001075F.p.24.1	M00005721D:B03
9	9/28/98	1492.001	9	RTA00001075F.o.04.1	M00005621B:C09
10	9/28/98	1492.001	10	RTA00000616F.j.04.1	M00005412D:G07
11	9/28/98	1492.001	11	RTA00001064F.k.01.1	M00005708C:D11
12	9/28/98	1492.001	12	RTA00001064F.j.19.1	M00005657B:F11
13	9/28/98	1492.001	13	RTA00001065F.a.22.1	M00006920B:H07
14	9/28/98	1492.001	14	RTA00001076F.d.11.1	M00006623C:G07
15	9/28/98	1492.001	15	RTA00000615F.e.08.2	M00004872A:D07
16	9/28/98	1492.001	16	RTA00000617F.p.05.2	M00005515D:G02
17	9/28/98	1492.001	17	RTA00001076F.f.03.1	M00006668D:B10
18	9/28/98	1492.001	18	RTA00001064F.l.17.2	M00006582A:F12
19	9/28/98	1492.001	19	RTA00001076F.h.13.1	M00006745B:C05
20	9/28/98	1492.001	20	RTA00001075F.k.12.1	M00005482A:D08
21	9/28/98	1492.001	21	RTA00001076F.c.09.1	M00006594B:D05
22	9/28/98	1492.001	22	RTA00001076F.l.16.1	M00006919A:H12
23	9/28/98	1492.001	23	RTA00001076F.b.13.1	M00005825A:A10
24	9/28/98	1492.001	24	RTA00001065F.d.06.2	M00007078B:H04
25	9/28/98	1492.001	25	RTA00001075F.p.23.1	M00005721C:A12
26	9/28/98	1492.001	26	RTA00001075F.n.22.1	M00005616B:E11
27	9/28/98	1492.001	27	RTA00001075F.o.21.1	M00005648C:E10
28	9/28/98	1492.001	28	RTA00001065F.b.22.1	M00006968A:H05
29	9/28/98	1492.001	29	RTA00001075F.p.06.1	M00005698A:H12
30	9/28/98	1492.001	30	RTA00001076F.d.19.1	M00006630A:E05
31	9/28/98	1492.001	31	RTA00001075F.e.14.1	M00005375B:H03
32	9/28/98	1492.001	32	RTA00001065F.f.02.1	M00007186A:A12
33	9/28/98	1492.001	33	RTA00001064F.p.03.1	M00006814D:D09
34	9/28/98	1492.001	34	RTA00001076F.i.19.1	M00006813B:E04
35	9/28/98	1492.001	35	RTA00001077F.c.06.1	M00007157B:B04
36	9/28/98	1492.001	36	RTA00001064F.c.21.1	M00005366D:E12
37	9/28/98	1492.001	37	RTA00001065F.e.21.1	M00007177A:G07
38	9/28/98	1492.001	38	RTA00001076F.o.14.1	M00007038D:D01

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
39	9/28/98	1492.001	39	RTA00001064F.c.01.1	M00005327C:G08
40	9/28/98	1492.001	40	RTA00001064F.d.16.1	M00005397A:G08
41	9/28/98	1492.001	41	RTA00000615F.e.05.2	M00004870D:E05
42	9/28/98	1492.001	42	RTA00000616F.j.12.1	M00005413D:G12
43	9/28/98	1492.001	43	RTA00001075F.a.17.1	M00004852B:H08
44	9/28/98	1492.001	44	RTA00001076F.n.10.1	M00006989C:B01
45	9/28/98	1492.001	45	RTA00001075F.l.04.1	M00005505D:H08
46	9/28/98	1492.001	46	RTA00001075F.l.10.1	M00005509B:E10
47	9/28/98	1492.001	47	RTA00001075F.i.09.1	M00005444D:D01
48	9/28/98	1492.001	48	RTA00001075F.j.13.1	M00005464B:B08
49	9/28/98	1492.001	49	RTA00001076F.e.03.1	M00006635A:C01
50	9/28/98	1492.001	50	RTA00001076F.j.14.1	M00006837B:H12
51	9/28/98	1492.001	51	RTA00001075F.g.19.1	M00005418C:B09
52	9/28/98	1492.001	52	RTA00001075F.m.05.1	M00005538C:H11
53	9/28/98	1492.001	53	RTA00001076F.p.03.1	M00007046D:E10
54	9/28/98	1492.001	54	RTA00001075F.h.19.1	M00005435B:F01
55	9/28/98	1492.001	55	RTA00001075F.h.14.1	M00005434C:E02
56	9/28/98	1492.001	56	RTA00001076F.l.14.1	M00006917B:C05
57	9/28/98	1492.001	57	RTA00001075F.h.17.1	M00005434D:H02
58	9/28/98	1492.001	58	RTA00001075F.f.18.1	M00005396C:H04
59	9/28/98	1492.001	59	RTA00001076F.l.03.1	M00006894D:A07
60	9/28/98	1492.001	60	RTA00001065F.d.07.2	M00007079D:H01
61	9/28/98	1492.001	61	RTA00001075F.e.18.1	M00005377C:F07
62	9/28/98	1492.001	62	RTA00001065F.d.03.2	M00007065D:A03
63	9/28/98	1492.001	63	RTA00001076F.b.18.1	M00006577A:B01
64	9/28/98	1492.001	64	RTA00001075F.m.16.1	M00005569B:E04
65	9/28/98	1492.001	65	RTA00001076F.d.13.1	M00006627C:C02
66	9/28/98	1492.001	66	RTA00001076F.i.16.1	M00006805D:H12
67	9/28/98	1492.001	67	RTA00001076F.p.10.1	M00007064B:E09
68	9/28/98	1492.001	68	RTA00001064F.p.14.1	M00006835D:C08
69	9/28/98	1492.001	69	RTA00001077F.b.04.1	M00007126D:H01
70	9/28/98	1492.001	70	RTA00001076F.d.04.1	M00006619A:G11
71	9/28/98	1492.001	71	RTA00001077F.a.22.1	M00007121D:A11
72	9/28/98	1492.001	72	RTA00001077F.c.19.1	M00007178D:A10
73	9/28/98	1492.001	73	RTA00001065F.f.06.1	M00007197D:D12
74	9/28/98	1492.001	74	RTA00000616F.f.11.3	M00005395D:D11
75	9/28/98	1492.001	75	RTA00001064F.l.13.2	M00006577B:F01
76	9/28/98	1492.001	76	RTA00001064F.o.08.1	M00006757D:H04

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
77	9/28/98	1492.001	77	RTA00001075F.o.03.1	M00005621A:B05
78	9/28/98	1492.001	78	RTA00001064F.l.23.2	M00006596D:H02
79	9/28/98	1492.001	79	RTA00001076F.e.01.1	M00006631D:G09
80	9/28/98	1492.001	80	RTA00001075F.j.22.1	M00005473C:F02
81	9/28/98	1492.001	81	RTA00001076F.h.16.1	M00006757A:C09
82	9/28/98	1492.001	82	RTA00001075F.j.08.1	M00005459B:A01
83	9/28/98	1492.001	83	RTA00001064F.o.19.1	M00006795C:B12
84	9/28/98	1492.001	84	RTA00001064F.o.07.1	M00006756D:G07
85	9/28/98	1492.001	85	RTA00001076F.i.09.1	M00006790D:F10
86	9/28/98	1492.001	86	RTA00001076F.i.22.1	M00006815D:D11
87	9/28/98	1492.001	87	RTA00001076F.c.21.1	M00006613C:C02
88	9/28/98	1492.001	88	RTA00001076F.j.19.1	M00006846A:B03
89	9/28/98	1492.001	89	RTA00001064F.o.13.1	M00006779D:F03
90	9/28/98	1492.001	90	RTA00001077F.a.06.1	M00007101C:H01
91	9/28/98	1492.001	91	RTA00001064F.n.01.1	M00006664A:C05
92	9/28/98	1492.001	92	RTA00001064F.c.12.1	M00005358A:H03
93	9/28/98	1492.001	93	RTA00001077F.d.07.1	M00007196D:D02
94	9/28/98	1492.001	94	RTA00001077F.c.18.1	M00007177B:C02
95	9/28/98	1492.001	95	RTA00001064F.g.12.1	M00005490B:B02
96	9/28/98	1492.001	96	RTA00001075F.b.07.1	M00004866C:H08
97	9/28/98	1492.001	97	RTA00000617F.p.03.2	M00005515B:B08
98	9/28/98	1492.001	98	RTA00000616F.f.10.3	M00005395D:B12
99	9/28/98	1492.001	99	RTA00001064F.p.15.1	M00006840A:A12
100	9/28/98	1492.001	100	RTA00000617F.p.10.2	M00005516D:F12
101	9/28/98	1492.001	101	RTA00001076F.m.01.1	M00006925B:B02
102	9/28/98	1492.001	102	RTA00001075F.f.15.1	M00005395C:C11
103	9/28/98	1492.001	103	RTA00001075F.e.23.1	M00005385B:A10
104	9/28/98	1492.001	104	RTA00001076F.f.12.1	M00006688C:C12
105	9/28/98	1492.001	105	RTA00001075F.g.21.1	M00005420C:E03
106	9/28/98	1492.001	106	RTA00001076F.g.18.1	M00006727A:H12
107	9/28/98	1492.001	107	RTA00001075F.d.24.1	M00005363D:C05
108	9/28/98	1492.001	108	RTA00001075F.e.02.1	M00005364C:A02
109	9/28/98	1492.001	109	RTA00001075F.m.14.1	M00005563C:D05
110	9/28/98	1492.001	110	RTA00001064F.h.07.1	M00005520A:H11
111	9/28/98	1492.001	111	RTA00001065F.b.07.1	M00006936C:G11
112	9/28/98	1492.001	112	RTA00001065F.b.23.1	M00006968D:H02
113	9/28/98	1492.001	113	RTA00001064F.g.15.1	M00005497C:G08
114	9/28/98	1492.001	114	RTA00001064F.d.14.1	M00005390C:E05

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
115	9/28/98	1492.001	115	RTA00001064F.l.22.2	M00006595C:B08
116	9/28/98	1492.001	116	RTA00001064F.p.04.1	M00006816D:D08
117	9/28/98	1492.001	117	RTA00001076F.g.04.1	M00006712A:F01
118	9/28/98	1492.001	118	RTA00001075F.p.17.1	M00005709D:H05
119	9/28/98	1492.001	119	RTA00001075F.l.03.1	M00005505B:D10
120	9/28/98	1492.001	120	RTA00001076F.l.23.1	M00006925A:B09
121	9/28/98	1492.001	121	RTA00001076F.k.11.1	M00006874D:E01
122	9/28/98	1492.001	122	RTA00001076F.n.15.1	M00006994A:C12
123	9/28/98	1492.001	123	RTA00001075F.o.10.1	M00005629B:G06
124	9/28/98	1492.001	124	RTA00001075F.n.04.1	M00005589B:H12
125	9/28/98	1492.001	125	RTA00001075F.f.06.1	M00005388B:B02
126	9/28/98	1492.001	126	RTA00001076F.j.05.1	M00006823A:H06
127	9/28/98	1492.001	127	RTA00001076F.o.18.1	M00007041C:C05
128	9/28/98	1492.001	128	RTA00001064F.j.14.1	M00005648C:C11
129	9/28/98	1492.001	129	RTA00001064F.d.06.1	M00005376B:E08
130	9/28/98	1492.001	130	RTA00001077F.d.10.1	M00007200A:B12
131	9/28/98	1492.001	131	RTA00001065F.d.19.1	M00007109D:G01
132	9/28/98	1492.001	132	RTA00001064F.f.13.1	M00005464D:D07
133	9/28/98	1492.001	133	RTA00001075F.k.20.1	M00005493D:H12
134	9/28/98	1492.001	134	RTA00001075F.k.07.1	M00005479C:A05
135	9/28/98	1492.001	135	RTA00001075F.a.14.1	M00004847D:G01
136	9/28/98	1492.001	136	RTA00001076F.f.22.1	M00006704A:C11
137	9/28/98	1492.001	137	RTA00001076F.m.11.1	M00006949B:C07
138	9/28/98	1492.001	138	RTA00001064F.i.13.2	M00005618C:H11
139	9/28/98	1492.001	139	RTA00001076F.f.19.3	M00006694D:G06
140	9/28/98	1492.001	140	RTA00001076F.c.23.1	M00006617A:A06
141	9/28/98	1492.001	141	RTA00001077F.a.09.1	M00007107C:D02
142	9/28/98	1492.001	142	RTA00001064F.b.14.1	M00005020B:D10
143	9/28/98	1492.001	143	RTA00001075F.e.21.1	M00005382A:G09
144	9/28/98	1492.001	144	RTA00001075F.p.15.1	M00005705D:G09
145	9/28/98	1492.001	145	RTA00001076F.n.11.1	M00006991B:E05
146	9/28/98	1492.001	146	RTA00001065F.e.18.1	M00007161C:D12
147	9/28/98	1492.001	147	RTA00000615F.e.06.2	M00004871C:C04
148	9/28/98	1492.001	148	RTA00001064F.a.04.2	M00004821D:C03
149	9/28/98	1492.001	149	RTA00001075F.j.18.1	M00005469A:D10
150	9/28/98	1492.001	150	RTA00001077F.c.05.1	M00007156D:E11
151	9/28/98	1492.001	151	RTA00001075F.g.22.1	M00005420C:E10
152	9/28/98	1492.001	152	RTA00001077F.a.08.1	M00007104D:D10

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
153	9/28/98	1492.001	153	RTA00001077F.c.15.1	M00007172D:H03
154	9/28/98	1492.001	154	RTA00001077F.c.16.1	M00007175B:B11
155	9/28/98	1492.001	155	RTA00001077F.b.15.1	M00007141A:G08
156	9/28/98	1492.001	156	RTA00001077F.c.17.1	M00007175D:G02
157	9/28/98	1492.001	157	RTA00001077F.a.14.1	M00007116A:C08
158	9/28/98	1492.001	158	RTA00001075F.i.02.1	M00005438D:A08
159	9/28/98	1492.001	159	RTA00001075F.l.11.1	M00005509D:G05
160	9/28/98	1492.001	160	RTA00001064F.d.20.1	M00005403A:D12
161	9/28/98	1492.001	161	RTA00001076F.h.10.1	M00006740A:A06
162	9/28/98	1492.001	162	RTA00001075F.k.21.1	M00005494C:F08
163	9/28/98	1492.001	163	RTA00001075F.i.21.1	M00005450C:G09
164	9/28/98	1492.001	164	RTA00001076F.p.24.1	M00007093C:C11
165	9/28/98	1492.001	165	RTA00001075F.f.03.1	M00005385D:B08
166	9/28/98	1492.001	166	RTA00001065F.d.18.2	M00007107A:H08
167	9/28/98	1492.001	167	RTA00001076F.o.05.1	M00007026A:A03
168	9/28/98	1492.001	168	RTA00001075F.d.10.1	M00005353C:H01
169	9/28/98	1492.001	169	RTA00001064F.d.07.1	M00005378B:B04
170	9/28/98	1492.001	170	RTA00001065F.b.11.1	M00006945D:A07
171	9/28/98	1492.001	171	RTA00001076F.g.17.1	M00006726D:H10
172	9/28/98	1492.001	172	RTA00001065F.a.21.1	M00006918D:G08
173	9/28/98	1492.001	173	RTA00001077F.d.12.1	M00007203C:E06
174	9/28/98	1492.001	174	RTA00001064F.g.08.1	M00005481C:H05
175	9/28/98	1492.001	175	RTA00001064F.f.02.1	M00005449D:D04
176	9/28/98	1492.001	176	RTA00001075F.a.02.1	M00004825A:G12
177	9/28/98	1492.001	177	RTA00001064F.b.16.1	M00005296B:H07
178	9/28/98	1492.001	178	RTA00001077F.c.02.1	M00007152A:A10
179	9/28/98	1492.001	179	RTA00001064F.g.04.1	M00005480C:A04
180	9/28/98	1492.001	180	RTA00001075F.c.12.1	M00005305A:H01
181	9/28/98	1492.001	181	RTA00001064F.o.04.1	M00006752C:D04
182	9/28/98	1492.001	182	RTA00001077F.a.21.1	M00007121A:G04
183	9/28/98	1492.001	183	RTA00001075F.f.11.1	M00005392C:B03
184	9/28/98	1492.001	184	RTA00001064F.k.24.2	M00005820A:H11
185	9/28/98	1492.001	185	RTA00001075F.d.02.1	M00005342D:E04
186	9/28/98	1492.001	186	RTA00001076F.c.13.1	M00006600D:G07
187	9/28/98	1492.001	187	RTA00001075F.b.15.1	M00004872C:G03
188	9/28/98	1492.001	188	RTA00001064F.f.09.1	M00005461C:D11
189	9/28/98	1492.001	189	RTA00001075F.g.14.1	M00005416B:A01
190	9/28/98	1492.001	190	RTA00001075F.f.17.1	M00005396A:C01

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
191	9/28/98	1492.001	191	RTA00001076F.l.05.1	M00006895D:A02
192	9/28/98	1492.001	192	RTA00001076F.o.02.1	M00007019B:G01
193	9/28/98	1492.001	193	RTA00001064F.b.07.1	M00005000A:H05
194	9/28/98	1492.001	194	RTA00001075F.d.17.1	M00005358B:D10
195	9/28/98	1492.001	195	RTA00000624F.f.12.2	M00005607A:C08
196	9/28/98	1492.001	196	RTA00001075F.c.22.1	M00005342B:G01
197	9/28/98	1492.001	197	RTA00001065F.a.17.1	M00006914C:D07
198	9/28/98	1492.001	198	RTA00001075F.b.02.1	M00004859D:D01
199	9/28/98	1492.001	199	RTA00001077F.c.12.1	M00007167C:B10
200	9/28/98	1492.001	200	RTA00001077F.c.20.1	M00007179B:H04
201	9/28/98	1492.001	201	RTA00001076F.m.04.1	M00006934B:B11
202	9/28/98	1492.001	202	RTA00001076F.j.22.1	M00006859D:E11
203	9/28/98	1492.001	203	RTA00001076F.k.13.1	M00006882C:D03
204	9/28/98	1492.001	204	RTA00001075F.k.14.1	M00005485C:F09
205	9/28/98	1492.001	205	RTA00001076F.f.10.1	M00006680D:A01
206	9/28/98	1492.001	206	RTA00001064F.o.05.1	M00006755C:C03
207	9/28/98	1492.001	207	RTA00001064F.l.05.2	M00005826B:F10
208	9/28/98	1492.001	208	RTA00001076F.p.04.1	M00007047D:C02
209	9/28/98	1492.001	209	RTA00001064F.l.04.1	M00005822D:C05
210	9/28/98	1492.001	210	RTA00001076F.c.03.1	M00006584D:D01
211	9/28/98	1492.001	211	RTA00001064F.m.06.1	M00006621B:B06
212	9/28/98	1492.001	212	RTA00001075F.k.15.1	M00005486A:F07
213	9/28/98	1492.001	213	RTA00001064F.d.08.1	M00005378C:B12
214	9/28/98	1492.001	214	RTA00001077F.d.11.1	M00007202A:A09
215	9/28/98	1492.001	215	RTA00001077F.b.14.1	M00007140C:G12
216	9/28/98	1492.001	216	RTA00001075F.k.04.1	M00005476D:A11
217	9/28/98	1492.001	217	RTA00001064F.n.03.1	M00006678C:B07
218	9/28/98	1492.001	218	RTA00001075F.i.12.1	M00005446B:D10
219	9/28/98	1492.001	219	RTA00001075F.f.04.1	M00005386C:G01
220	9/28/98	1492.001	220	RTA00001076F.n.14.1	M00006993B:F02
221	9/28/98	1492.001	221	RTA00001064F.k.19.2	M00005810B:C07
222	9/28/98	1492.001	222	RTA00001076F.d.20.1	M00006630A:E09
223	9/28/98	1492.001	223	RTA00001077F.b.20.1	M00007145C:B05
224	9/28/98	1492.001	224	RTA00001076F.f.11.1	M00006688A:F09
225	9/28/98	1492.001	225	RTA00001065F.d.01.1	M00007047C:H04
226	9/28/98	1492.001	226	RTA00001075F.g.12.1	M00005413B:B02
227	9/28/98	1492.001	227	RTA00001064F.a.09.2	M00004841C:H03
228	9/28/98	1492.001	228	RTA00001064F.k.20.2	M00005810B:G02

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
229	9/28/98	1492.001	229	RTA00001064F.b.17.1	M00005296D:G03
230	9/28/98	1493.001	1	RTA00001073F.f.17.1	M00004087A:H06
231	9/28/98	1493.001	2	RTA00001073F.l.02.1	M00004168D:F05
232	9/28/98	1493.001	3	RTA00001072F.i.07.3	M00003845B:A04
233	9/28/98	1493.001	4	RTA00001071F.i.23.3	M00001477A:G02
234	9/28/98	1493.001	5	RTA00000611F.e.04.2	M00004170C:H06
235	9/28/98	1493.001	6	RTA00001062F.f.19.1	M00003888C:G08
236	9/28/98	1493.001	7	RTA00001073F.l.22.1	M00004176B:H09
237	9/28/98	1493.001	8	RTA00001063F.l.10.1	M00004410A:F06
238	9/28/98	1493.001	9	RTA00001062F.l.13.1	M00004034A:A05
239	9/28/98	1493.001	10	RTA00001074F.l.10.1	M00004495D:A05
240	9/28/98	1493.001	11	RTA00001061F.d.01.1	M00001389C:E01
241	9/28/98	1493.001	12	RTA00001072F.j.04.2	M00003861D:G10
242	9/28/98	1493.001	13	RTA00001073F.d.04.1	M00004048C:C02
243	9/28/98	1493.001	14	RTA00001061F.j.09.1	M00001507A:H06
244	9/28/98	1493.001	15	RTA00001071F.h.16.1	M00001450D:H12
245	9/28/98	1493.001	16	RTA00001062F.o.17.1	M00004108B:D04
246	9/28/98	1493.001	17	RTA00001073F.c.20.1	M00004046C:A04
247	9/28/98	1493.001	18	RTA00001063F.k.14.1	M00004381A:E10
248	9/28/98	1493.001	19	RTA00000611F.e.18.2	M00004171D:H10
249	9/28/98	1493.001	20	RTA00001072F.a.18.2	M00001655C:F07
250	9/28/98	1493.001	21	RTA00001072F.b.04.2	M00001660A:B10
251	9/28/98	1493.001	22	RTA00001074F.g.19.1	M00004372A:A08
252	9/28/98	1493.001	23	RTA00001072F.i.09.3	M00003845C:F08
253	9/28/98	1493.001	24	RTA00001072F.a.21.2	M00001657D:D07
254	9/28/98	1493.001	25	RTA00001072F.m.18.3	M00003916D:A10
255	9/28/98	1493.001	26	RTA00001061F.b.04.1	M00001360B:F09
256	9/28/98	1493.001	27	RTA00001072F.o.06.2	M00003935A:C04
257	9/28/98	1493.001	28	RTA00001072F.n.19.3	M00003931A:G01
258	9/28/98	1493.001	29	RTA00001073F.e.08.1	M00004068A:A03
259	9/28/98	1493.001	30	RTA00001074F.g.22.1	M00004373D:G10
260	9/28/98	1493.001	31	RTA00001073F.c.01.1	M00004030C:E05
261	9/28/98	1493.001	32	RTA00001074F.f.15.1	M00004360B:B08
262	9/28/98	1493.001	33	RTA00001074F.f.01.1	M00004350A:C04
263	9/28/98	1493.001	34	RTA00001074F.d.08.1	M00004318D:D07
264	9/28/98	1493.001	35	RTA00001072F.f.11.2	M00003788D:E06
265	9/28/98	1493.001	36	RTA00001074F.e.05.1	M00004337A:A07
266	9/28/98	1493.001	37	RTA00001072F.g.05.2	M00003803B:G12

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
267	9/28/98	1493.001	38	RTA00001071F.j.04.3	M00001479D:B10
268	9/28/98	1493.001	39	RTA00001074F.j.05.1	M00004415A:A01
269	9/28/98	1493.001	40	RTA00001074F.j.04.1	M00004414D:C11
270	9/28/98	1493.001	41	RTA00001073F.e.06.1	M00004067C:C10
271	9/28/98	1493.001	42	RTA00001071F.d.14.1	M00001389A:F03
272	9/28/98	1493.001	43	RTA00001071F.f.12.1	M00001418C:F06
273	9/28/98	1493.001	44	RTA00001061F.m.13.1	M00001601D:A03
274	9/28/98	1493.001	45	RTA00001061F.e.17.1	M00001418A:A02
275	9/28/98	1493.001	46	RTA00001071F.m.09.3	M00001563A:F04
276	9/28/98	1493.001	47	RTA00001062F.l.05.1	M00004029D:H03
277	9/28/98	1493.001	48	RTA00001073F.i.02.2	M00004125B:A02
278	9/28/98	1493.001	49	RTA00001063F.l.04.1	M00004404C:B03
279	9/28/98	1493.001	50	RTA00001063F.l.14.1	M00004412A:G05
280	9/28/98	1493.001	51	RTA00001063F.e.05.1	M00004232D:G11
281	9/28/98	1493.001	52	RTA00001062F.f.06.1	M00003880A:G10
282	9/28/98	1493.001	53	RTA00001072F.b.23.2	M00001683B:F12
283	9/28/98	1493.001	54	RTA00001073F.a.13.1	M00003989D:A02
284	9/28/98	1493.001	55	RTA00001074F.h.16.1	M00004386C:C03
285	9/28/98	1493.001	56	RTA00001073F.a.15.1	M00003991A:D05
286	9/28/98	1493.001	57	RTA00001073F.k.01.1	M00004152A:F03
287	9/28/98	1493.001	58	RTA00001072F.l.19.2	M00003901B:C02
288	9/28/98	1493.001	59	RTA00001072F.i.15.3	M00003848A:E08
289	9/28/98	1493.001	60	RTA00001072F.i.05.3	M00003844D:B02
290	9/28/98	1493.001	61	RTA00001074F.m.06.1	M00004603D:D09
291	9/28/98	1493.001	62	RTA00001062F.m.15.1	M00004063B:B12
292	9/28/98	1493.001	63	RTA00001074F.d.19.1	M00004326D:D06
293	9/28/98	1493.001	64	RTA00001073F.j.02.1	M00004140B:C02
294	9/28/98	1493.001	65	RTA00001071F.l.11.1	M00001545D:F12
295	9/28/98	1493.001	66	RTA00001074F.f.12.1	M00004356C:D02
296	9/28/98	1493.001	67	RTA00001073F.h.03.1	M00004110A:G03
297	9/28/98	1493.001	68	RTA00001074F.a.19.1	M00004275A:H07
298	9/28/98	1493.001	69	RTA00001063F.g.15.1	M00004292A:C08
299	9/28/98	1493.001	70	RTA00001061F.a.09.1	M00001345C:B10
300	9/28/98	1493.001	71	RTA00001063F.f.23.1	M00004284A:C09
301	9/28/98	1493.001	72	RTA00001073F.e.10.1	M00004069A:E04
302	9/28/98	1493.001	73	RTA00001073F.g.15.1	M00004103A:E06
303	9/28/98	1493.001	74	RTA00001073F.n.20.1	M00004209B:G01
304	9/28/98	1493.001	75	RTA00001073F.g.11.1	M00004099C:F04

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
305	9/28/98	1493.001	76	RTA00001071F.p.05.1	M00001630A:E08
306	9/28/98	1493.001	77	RTA00001073F.l.19.1	M00004175D:D05
307	9/28/98	1493.001	78	RTA00001074F.j.17.1	M00004426B:H06
308	9/28/98	1493.001	79	RTA00001074F.b.22.1	M00004292A:F03
309	9/28/98	1493.001	80	RTA00001071F.d.19.1	M00001391C:B05
310	9/28/98	1493.001	81	RTA00001062F.j.02.1	M00003960D:E09
311	9/28/98	1493.001	82	RTA00001072F.b.09.2	M00001664D:E02
312	9/28/98	1493.001	83	RTA00001073F.b.08.1	M00003998C:D04
313	9/28/98	1493.001	84	RTA00001062F.j.19.1	M00003977D:H04
314	9/28/98	1493.001	85	RTA00001062F.m.18.1	M00004066D:C02
315	9/28/98	1493.001	86	RTA00001062F.b.02.1	M00003775C:C01
316	9/28/98	1493.001	87	RTA00001061F.d.20.1	M00001401B:A02
317	9/28/98	1493.001	88	RTA00001071F.n.05.3	M00001579C:E07
318	9/28/98	1493.001	89	RTA00001073F.l.04.1	M00004170B:G04
319	9/28/98	1493.001	90	RTA00001071F.h.04.1	M00001442D:D09
320	9/28/98	1493.001	91	RTA00001062F.o.11.1	M00004104C:F06
321	9/28/98	1493.001	92	RTA00001062F.i.10.1	M00003939B:C02
322	9/28/98	1493.001	93	RTA00001071F.g.16.1	M00001431A:F03
323	9/28/98	1493.001	94	RTA00001061F.d.06.1	M00001392A:F02
324	9/28/98	1493.001	95	RTA00001071F.m.01.3	M00001561A:G10
325	9/28/98	1493.001	96	RTA00001062F.n.06.1	M00004081A:E11
326	9/28/98	1493.001	97	RTA00001061F.d.14.1	M00001397D:G04
327	9/28/98	1493.001	98	RTA00001061F.j.10.1	M00001507D:F09
328	9/28/98	1493.001	99	RTA00001063F.c.07.1	M00004185B:H03
329	9/28/98	1493.001	100	RTA00001061F.j.12.1	M00001513B:F05
330	9/28/98	1493.001	101	RTA00001061F.o.22.1	M00001678A:B10
331	9/28/98	1493.001	102	RTA00001071F.e.03.1	M00001395D:B04
332	9/28/98	1493.001	103	RTA00001072F.e.13.2	M00003772C:F12
333	9/28/98	1493.001	104	RTA00001062F.i.03.1	M00003928D:A04
334	9/28/98	1493.001	105	RTA00001072F.d.20.2	M00003761C:C05
335	9/28/98	1493.001	106	RTA00001074F.g.16.1	M00004371B:A05
336	9/28/98	1493.001	107	RTA00001074F.f.09.1	M00004353D:C06
337	9/28/98	1493.001	108	RTA00001071F.k.12.1	M00001505C:C10
338	9/28/98	1493.001	109	RTA00001074F.f.13.1	M00004357A:B10
339	9/28/98	1493.001	110	RTA00001071F.e.08.1	M00001397C:F01
340	9/28/98	1493.001	111	RTA00001073F.h.11.1	M00004117D:F06
341	9/28/98	1493.001	112	RTA00001072F.o.14.2	M00003937D:F09
342	9/28/98	1493.001	113	RTA00001074F.c.11.1	M00004298A:H09

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
343	9/28/98	1493.001	114	RTA00001074F.g.08.1	M00004368A:G11
344	9/28/98	1493.001	115	RTA00001073F.a.18.1	M00003993C:G11
345	9/28/98	1493.001	116	RTA00001073F.f.19.1	M00004090A:B11
346	9/28/98	1493.001	117	RTA00001072F.l.20.2	M00003902C:D02
347	9/28/98	1493.001	118	RTA00001073F.b.06.1	M00003997D:G03
348	9/28/98	1493.001	119	RTA00001062F.o.14.1	M00004105C:C05
349	9/28/98	1493.001	120	RTA00001071F.i.04.3	M00001457D:E08
350	9/28/98	1493.001	121	RTA00001074F.a.23.1	M00004278C:H11
351	9/28/98	1493.001	122	RTA00001073F.c.04.1	M00004034A:G03
352	9/28/98	1493.001	123	RTA00001072F.h.18.2	M00003833D:F11
353	9/28/98	1493.001	124	RTA00001074F.i.06.1	M00004403A:A02
354	9/28/98	1493.001	125	RTA00001063F.e.09.1	M00004240A:D03
355	9/28/98	1493.001	126	RTA00001061F.d.03.1	M00001390C:H05
356	9/28/98	1493.001	127	RTA00001063F.d.23.1	M00004225A:E03
357	9/28/98	1493.001	128	RTA00001063F.k.08.1	M00004378A:H10
358	9/28/98	1493.001	129	RTA00001062F.b.04.1	M00003776B:F08
359	9/28/98	1493.001	130	RTA00001063F.b.18.1	M00004178B:F07
360	9/28/98	1493.001	131	RTA00001062F.b.11.1	M00003788B:C08
361	9/28/98	1493.001	132	RTA00001074F.l.23.1	M00004504C:G07
362	9/28/98	1493.001	133	RTA00001063F.m.08.1	M00004444C:H11
363	9/28/98	1493.001	134	RTA00001071F.l.13.2	M00001549C:F10
364	9/28/98	1493.001	135	RTA00001072F.p.19.2	M00003973A:D09
365	9/28/98	1493.001	136	RTA00001071F.k.17.1	M00001517C:A10
366	9/28/98	1493.001	137	RTA00001072F.o.24.2	M00003943B:C12
367	9/28/98	1493.001	138	RTA00001074F.a.20.1	M00004276A:C06
368	9/28/98	1493.001	139	RTA00001073F.c.16.1	M00004043C:A06
369	9/28/98	1493.001	140	RTA00001074F.j.10.1	M00004422C:A01
370	9/28/98	1493.001	141	RTA00001063F.n.16.1	M00004498D:F02
371	9/28/98	1493.001	142	RTA00001071F.o.16.1	M00001615A:D01
372	9/28/98	1493.001	143	RTA00001073F.k.16.1	M00004165D:H12
373	9/28/98	1493.001	144	RTA00001062F.e.14.1	M00003856A:H10
374	9/28/98	1493.001	145	RTA00001071F.h.22.1	M00001454D:H09
375	9/28/98	1493.001	146	RTA00001071F.o.18.1	M00001618C:E01
376	9/28/98	1493.001	147	RTA00001062F.p.19.1	M00004140D:E03
377	9/28/98	1493.001	148	RTA00001062F.d.04.1	M00003818C:D02
378	9/28/98	1493.001	149	RTA00001072F.n.22.3	M00003933A:B04
379	9/28/98	1493.001	150	RTA00001063F.c.11.1	M00004187A:B05
380	9/28/98	1493.001	151	RTA00001061F.j.22.1	M00001531B:A03

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
381	9/28/98	1493.001	152	RTA00001062F.d.08.1	M00003820C:E08
382	9/28/98	1493.001	153	RTA00001062F.f.02.1	M00003877C:G01
383	9/28/98	1493.001	154	RTA00001062F.d.24.1	M00003839D:C03
384	9/28/98	1493.001	155	RTA00001074F.h.24.1	M00004391C:F12
385	9/28/98	1493.001	156	RTA00001071F.a.10.1	M00001341A:H10
386	9/28/98	1493.001	157	RTA00001074F.k.13.1	M00004449B:B05
387	9/28/98	1493.001	158	RTA00001072F.k.16.2	M00003884C:G09
388	9/28/98	1493.001	159	RTA00001073F.k.09.1	M00004158C:B01
389	9/28/98	1493.001	160	RTA00001074F.b.14.1	M00004288D:E07
390	9/28/98	1493.001	161	RTA00001073F.k.08.1	M00004157C:E06
391	9/28/98	1493.001	162	RTA00001074F.i.17.1	M00004406D:E11
392	9/28/98	1493.001	163	RTA00001074F.k.10.1	M00004447A:A10
393	9/28/98	1493.001	164	RTA00001062F.p.14.1	M00004135D:D01
394	9/28/98	1493.001	165	RTA00001071F.m.15.3	M00001569A:H01
395	9/28/98	1493.001	166	RTA00001074F.h.15.1	M00004385D:D06
396	9/28/98	1493.001	167	RTA00001062F.i.09.1	M00003935D:E04
397	9/28/98	1493.001	168	RTA00000611F.e.06.2	M00004170D:C06
398	9/28/98	1493.001	169	RTA00001062F.d.19.1	M00003835B:C05
399	9/28/98	1493.001	170	RTA00001062F.o.15.1	M00004107A:E02
400	9/28/98	1493.001	171	RTA00001071F.a.07.1	M00001340C:A08
401	9/28/98	1493.001	172	RTA00001062F.d.07.1	M00003820B:G04
402	9/28/98	1493.001	173	RTA00001074F.j.11.1	M00004423A:B05
403	9/28/98	1493.001	174	RTA00001071F.m.11.3	M00001565C:F06
404	9/28/98	1493.001	175	RTA00001062F.i.01.1	M00003926A:D01
405	9/28/98	1493.001	176	RTA00001072F.g.08.2	M00003804D:F12
406	9/28/98	1493.001	177	RTA00001071F.n.16.1	M00001594A:H01
407	9/28/98	1493.001	178	RTA00001062F.a.09.1	M00003756D:B09
408	9/28/98	1493.001	179	RTA00001073F.h.08.1	M00004114C:B09
409	9/28/98	1493.001	180	RTA00001073F.e.03.1	M00004064B:G03
410	9/28/98	1493.001	181	RTA00001073F.c.23.1	M00004048A:E10
411	9/28/98	1493.001	182	RTA00001074F.l.15.1	M00004498D:A11
412	9/28/98	1493.001	183	RTA00001073F.l.21.1	M00004176A:H05
413	9/28/98	1493.001	184	RTA00001071F.d.15.1	M00001389B:B12
414	9/28/98	1493.001	185	RTA00001073F.i.08.1	M00004127C:C08
415	9/28/98	1493.001	186	RTA00001073F.k.21.1	M00004167A:H04
416	9/28/98	1493.001	187	RTA00001072F.j.05.2	M00003865B:D10
417	9/28/98	1493.001	188	RTA00001063F.i.15.1	M00004335A:G05
418	9/28/98	1493.001	189	RTA00001062F.g.21.1	M00003907C:D02

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
419	9/28/98	1493.001	190	RTA00001073F.b.16.1	M00004027C:E06
420	9/28/98	1493.001	191	RTA00001062F.g.06.1	M00003895C:F05
421	9/28/98	1493.001	192	RTA00001071F.b.17.1	M00001360B:B01
422	9/28/98	1493.001	193	RTA00001073F.f.18.1	M00004087B:D05
423	9/28/98	1493.001	194	RTA00001074F.b.04.1	M00004280D:D10
424	9/28/98	1493.001	195	RTA00001072F.d.23.2	M00003762D:C02
425	9/28/98	1493.001	196	RTA00001073F.l.14.1	M00004173A:D03
426	9/28/98	1493.001	197	RTA00001061F.p.21.1	M00003747C:G12
427	9/28/98	1493.001	198	RTA00001071F.n.22.1	M00001598C:F02
428	9/28/98	1493.001	199	RTA00001073F.d.22.1	M00004059D:A09
429	9/28/98	1493.001	200	RTA00001072F.j.14.2	M00003876C:G11
430	9/28/98	1493.001	201	RTA00001071F.k.21.2	M00001528D:B12
431	9/28/98	1493.001	202	RTA00001074F.a.09.1	M00004269C:B10
432	9/28/98	1493.001	203	RTA00001073F.p.19.1	M00004253A:E02
433	9/28/98	1493.001	204	RTA00001061F.b.02.1	M00001358B:F12
434	9/28/98	1493.001	205	RTA00001063F.e.10.1	M00004240C:A06
435	9/28/98	1493.001	206	RTA00001074F.j.18.1	M00004427D:H04
436	9/28/98	1493.001	207	RTA00001073F.f.09.1	M00004084C:F05
437	9/28/98	1493.001	208	RTA00001071F.l.19.1	M00001558D:E02
438	9/28/98	1493.001	209	RTA00001073F.c.09.1	M00004036B:C11
439	9/28/98	1493.001	210	RTA00001074F.a.14.1	M00004270C:H05
440	9/28/98	1493.001	211	RTA00001074F.l.03.1	M00004466A:E04
441	9/28/98	1493.001	212	RTA00000611F.f.13.2	M00004175D:G10
442	9/28/98	1493.001	213	RTA00001074F.e.16.1	M00004343A:G07
443	9/28/98	1493.001	214	RTA00001073F.l.05.1	M00004170C:A12
444	9/28/98	1493.001	215	RTA00001074F.e.19.1	M00004347A:F10
445	9/28/98	1493.001	216	RTA00001073F.e.07.1	M00004067C:E05
446	9/28/98	1493.001	217	RTA00001062F.p.22.1	M00004142C:A06
447	9/28/98	1493.001	218	RTA00001061F.c.11.1	M00001382D:F03
448	9/28/98	1493.001	219	RTA00001062F.f.01.1	M00003877C:A08
449	9/28/98	1493.001	220	RTA00001072F.l.09.2	M00003893A:D03
450	9/28/98	1493.001	221	RTA00001072F.i.14.2	M00003847B:H01
451	9/28/98	1493.001	222	RTA00001063F.g.18.1	M00004295A:C02
452	9/28/98	1493.001	223	RTA00001062F.j.18.1	M00003977C:D01
453	9/28/98	1493.001	224	RTA00001061F.b.05.1	M00001360D:C12
454	9/28/98	1493.001	225	RTA00001074F.e.18.1	M00004344B:C06
455	9/28/98	1493.001	226	RTA00001061F.o.20.1	M00001677B:G01
456	9/28/98	1493.001	227	RTA00001062F.d.10.1	M00003822A:D02

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
457	9/28/98	1493.001	228	RTA00001062F.h.16.1	M00003919D:F01
458	9/28/98	1493.001	229	RTA00001063F.e.19.1	M00004251B:H12
459	9/28/98	1493.001	230	RTA00001061F.o.18.1	M00001675C:F05
460	9/28/98	1493.001	231	RTA00001072F.j.20.2	M00003879D:A09
461	9/28/98	1493.001	232	RTA00001071F.j.15.3	M00001485A:C04
462	9/28/98	1493.001	233	RTA00001071F.a.09.1	M00001340C:D09
463	9/28/98	1493.001	234	RTA00001074F.j.13.1	M00004423C:F03
464	9/28/98	1493.001	235	RTA00001071F.i.15.3	M00001466C:H11
465	9/28/98	1493.001	236	RTA00001071F.b.13.1	M00001358C:D09
466	9/28/98	1493.001	237	RTA00001061F.g.05.1	M00001441D:G02
467	9/28/98	1493.001	238	RTA00001063F.e.16.1	M00004249A:C09
468	9/28/98	1493.001	239	RTA00001072F.j.22.2	M00003880B:B08
469	9/28/98	1493.001	240	RTA00001063F.i.16.1	M00004335D:D03
470	9/28/98	1493.001	241	RTA00000611F.f.05.2	M00004174B:B12
471	9/28/98	1493.001	242	RTA00001071F.p.07.1	M00001631D:G08
472	9/28/98	1493.001	243	RTA00001071F.c.12.1	M00001375C:C11
473	9/28/98	1493.001	244	RTA00001074F.k.15.1	M00004450A:G07
474	9/28/98	1493.001	245	RTA00001061F.e.19.1	M00001419A:E01
475	9/28/98	1493.001	246	RTA00001073F.g.22.1	M00004108C:D07
476	9/28/98	1493.001	247	RTA00001061F.g.01.1	M00001437D:A12
477	9/28/98	1493.001	248	RTA00001072F.n.08.2	M00003923D:A03
478	9/28/98	1493.001	249	RTA00001074F.b.12.1	M00004286D:D02
479	9/28/98	1493.001	250	RTA00001061F.l.18.1	M00001576C:E03
480	9/28/98	1493.001	251	RTA00001074F.j.03.1	M00004414D:A01
481	9/28/98	1493.001	252	RTA00001072F.h.07.2	M00003824A:B11
482	9/28/98	1493.001	253	RTA00001072F.j.18.2	M00003877C:C11
483	9/28/98	1493.001	254	RTA00001063F.c.21.1	M00004198B:G08
484	9/28/98	1493.001	255	RTA00001073F.m.11.1	M00004181A:B05
485	9/28/98	1493.001	256	RTA00001061F.h.16.1	M00001463C:E12
486	9/28/98	1493.001	257	RTA00001073F.i.11.1	M00004128B:H11
487	9/28/98	1493.001	258	RTA00001062F.k.20.1	M00003997A:C08
488	9/28/98	1493.001	259	RTA00001062F.o.05.1	M00004101A:C12
489	9/28/98	1493.001	260	RTA00001073F.p.01.1	M00004237B:G01
490	9/28/98	1493.001	261	RTA00001072F.a.04.2	M00001647D:A02
491	9/28/98	1493.001	262	RTA00001073F.e.12.1	M00004071C:B06
492	9/28/98	1493.001	263	RTA00001073F.p.22.1	M00004253D:D04
493	9/28/98	1493.001	264	RTA00001072F.i.19.3	M00003853C:A09
494	9/28/98	1493.001	265	RTA00001071F.d.06.1	M00001386B:E01

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
495	9/28/98	1493.001	266	RTA00001073F.j.20.1	M00004149C:D11
496	9/28/98	1493.001	267	RTA00001074F.l.20.1	M00004502B:G05
497	9/28/98	1493.001	268	RTA00001072F.h.14.2	M00003829C:G07
498	9/28/98	1493.001	269	RTA00001062F.b.13.1	M00003788C:C05
499	9/28/98	1493.001	270	RTA00001061F.j.14.1	M00001514B:C02
500	9/28/98	1493.001	271	RTA00001072F.j.11.2	M00003870C:H03
501	9/28/98	1493.001	272	RTA00001074F.m.01.1	M00004507A:F11
502	9/28/98	1493.001	273	RTA00001063F.f.03.1	M00004264B:F03
503	9/28/98	1493.001	274	RTA00001071F.l.21.1	M00001559D:E02
504	9/28/98	1493.001	275	RTA00001072F.b.11.2	M00001669B:H04
505	9/28/98	1493.001	276	RTA00001074F.i.16.1	M00004406A:H12
506	9/28/98	1493.001	277	RTA00001061F.j.03.1	M00001500A:A02
507	9/28/98	1493.001	278	RTA00001062F.n.16.1	M00004085B:D12
508	9/28/98	1493.001	279	RTA00001073F.j.03.1	M00004140C:D04
509	9/28/98	1493.001	280	RTA00001072F.k.01.2	M00003880C:D06
510	9/28/98	1493.001	281	RTA00001074F.k.08.1	M00004445D:A04
511	9/28/98	1493.001	282	RTA00001062F.k.05.1	M00003985B:F06
512	9/28/98	1493.001	283	RTA00001073F.h.01.1	M00004109A:B07
513	9/28/98	1493.001	284	RTA00000611F.f.15.2	M00004176A:E07
514	9/28/98	1493.001	285	RTA00001073F.b.01.1	M00003995B:C06
515	9/28/98	1493.001	286	RTA00001072F.c.16.2	M00001694B:H12
516	9/28/98	1493.001	287	RTA00001073F.c.10.1	M00004036C:E10
517	9/28/98	1493.001	288	RTA00001062F.g.22.1	M00003908C:C04
518	9/28/98	1493.001	289	RTA00001074F.d.15.1	M00004323B:G12
519	9/28/98	1493.001	290	RTA00001061F.c.12.1	M00001383C:C04
520	9/28/98	1493.001	291	RTA00001073F.k.15.1	M00004165B:E03
521	9/28/98	1493.001	292	RTA00001072F.j.23.2	M00003880B:D03
522	9/28/98	1493.001	293	RTA00001073F.j.21.1	M00004150A:B09
523	9/28/98	1493.001	294	RTA00001073F.h.20.1	M00004123B:G05
524	9/28/98	1493.001	295	RTA00001063F.g.05.1	M00004285C:B06
525	9/28/98	1493.001	296	RTA00001061F.a.21.1	M00001352D:A09
526	9/28/98	1493.001	297	RTA00001061F.d.17.1	M00001399B:C04
527	9/28/98	1493.001	298	RTA00001072F.h.04.2	M00003819D:B02
528	9/29/98	1494.001	1	RTA00001082F.j.11.1	M00027137D:F05
529	9/29/98	1494.001	2	RTA00001082F.h.08.1	M00027042D:E02
530	9/29/98	1494.001	3	RTA00001082F.e.15.1	M00026936D:D01
531	9/29/98	1494.001	4	RTA00001082F.l.21.1	M00027204B:A08
532	9/29/98	1494.001	5	RTA00001082F.e.05.1	M00026910C:C05

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
533	9/29/98	1494.001	6	RTA00001082F.i.07.1	M00027085C:H12
534	9/29/98	1494.001	7	RTA00001082F.i.12.1	M00027096B:A01
535	9/29/98	1494.001	8	RTA00001082F.m.12.1	M00027218C:D06
536	9/29/98	1494.001	9	RTA00001082F.p.16.1	M00027364D:E08
537	9/29/98	1494.001	10	RTA00001082F.g.22.1	M00027028B:C12
538	9/29/98	1494.001	11	RTA00001069F.e.20.1	M00026857A:F02
539	9/29/98	1494.001	12	RTA00001082F.c.05.3	M00026811A:H01
540	9/29/98	1494.001	13	RTA00001083F.c.15.1	M00027529B:B11
541	9/29/98	1494.001	14	RTA00001082F.f.08.1	M00026964C:H02
542	9/29/98	1494.001	15	RTA00001082F.o.01.1	M00027280D:H01
543	9/29/98	1494.001	16	RTA00001082F.l.05.1	M00027190B:F06
544	9/29/98	1494.001	17	RTA00001082F.l.10.1	M00027196A:A10
545	9/29/98	1494.001	18	RTA00001069F.i.06.1	M00026972A:F04
546	9/29/98	1494.001	19	RTA00001082F.o.21.1	M00027339D:E10
547	9/29/98	1494.001	20	RTA00001069F.c.13.1	M00023390A:C04
548	9/29/98	1494.001	21	RTA00001069F.g.11.1	M00026914C:H10
549	9/29/98	1494.001	22	RTA00001082F.e.21.1	M00026945B:C10
550	9/29/98	1494.001	23	RTA00001083F.a.18.1	M00027396C:B06
551	9/29/98	1494.001	24	RTA00001069F.a.21.1	M00023298B:G07
552	9/29/98	1494.001	25	RTA00001083F.a.17.1	M00027393D:F01
553	9/29/98	1494.001	26	RTA00001083F.a.23.1	M00027439B:A09
554	9/29/98	1494.001	27	RTA00001083F.e.18.1	M00027642C:D11
555	9/29/98	1494.001	28	RTA00001083F.e.04.1	M00027618A:B08
556	9/29/98	1494.001	29	RTA00001069F.j.21.1	M00027067A:B02
557	9/29/98	1494.001	30	RTA00001082F.h.20.1	M00027069D:F02
558	9/29/98	1494.001	31	RTA00001069F.o.03.1	M00027386D:C02
559	9/29/98	1494.001	32	RTA00001082F.l.04.1	M00027189C:D04
560	9/29/98	1494.001	33	RTA00001082F.o.05.1	M00027282D:G01
561	9/29/98	1494.001	34	RTA00001069F.a.11.1	M00023284B:G06
562	9/29/98	1494.001	35	RTA00001069F.n.05.1	M00027283C:H12
563	9/29/98	1494.001	36	RTA00001069F.a.22.1	M00023299B:A01
564	9/29/98	1494.001	37	RTA00001069F.h.10.1	M00026942C:A06
565	9/29/98	1494.001	38	RTA00001082F.h.19.1	M00027067B:E09
566	9/29/98	1494.001	39	RTA00001082F.b.05.1	M00023343B:C08
567	9/29/98	1494.001	40	RTA00001082F.j.05.1	M00027131C:E07
568	9/29/98	1494.001	41	RTA00001083F.b.09.1	M00027459A:G12
569	9/29/98	1494.001	42	RTA00001082F.d.07.3	M00026871C:F12
570	9/29/98	1494.001	43	RTA00001083F.c.03.1	M00027499B:G02

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
571	9/29/98	1494.001	44	RTA00001082F.f.01.1	M00026949A:F04
572	9/29/98	1494.001	45	RTA00001082F.h.12.1	M00027053C:B06
573	9/29/98	1494.001	46	RTA00001082F.a.03.1	M00023282B:H09
574	9/29/98	1494.001	47	RTA00001082F.l.03.1	M00027188A:D12
575	9/29/98	1494.001	48	RTA00001082F.k.04.1	M00027154B:D05
576	9/29/98	1494.001	49	RTA00001069F.b.18.1	M00023340A:A10
577	9/29/98	1494.001	50	RTA00001069F.o.21.1	M00027546B:A11
578	9/29/98	1494.001	51	RTA00001082F.k.01.1	M00027152D:H06
579	9/29/98	1494.001	52	RTA00001083F.a.14.1	M00027388A:G05
580	9/29/98	1494.001	53	RTA00001069F.k.01.1	M00027085A:G10
581	9/29/98	1494.001	54	RTA00001069F.h.09.1	M00026941C:E11
582	9/29/98	1494.001	55	RTA00001069F.o.11.1	M00027462D:A12
583	9/29/98	1494.001	56	RTA00001083F.a.22.1	M00027438D:A03
584	9/29/98	1494.001	57	RTA00001082F.m.21.1	M00027231C:D08
585	9/29/98	1494.001	58	RTA00001083F.f.18.1	M00027752B:E05
586	9/29/98	1494.001	59	RTA00001082F.i.03.1	M00027083C:F06
587	9/29/98	1494.001	60	RTA00001082F.n.01.1	M00027234C:B05
588	9/29/98	1494.001	61	RTA00001082F.l.02.1	M00027184D:H02
589	9/29/98	1494.001	62	RTA00001082F.k.18.1	M00027178B:E04
590	9/29/98	1494.001	63	RTA00001069F.d.09.1	M00023413D:F04
591	9/29/98	1494.001	64	RTA00001069F.p.05.1	M00027607A:A09
592	9/29/98	1494.001	65	RTA00001069F.m.14.1	M00027231A:D01
593	9/29/98	1494.001	66	RTA00001083F.c.21.1	M00027557D:B06
594	9/29/98	1494.001	67	RTA00001069F.i.23.1	M00027023B:H12
595	9/29/98	1494.001	68	RTA00001082F.l.07.1	M00027193A:F07
596	9/29/98	1494.001	69	RTA00001082F.c.15.3	M00026850B:F07
597	9/29/98	1494.001	70	RTA00001082F.f.18.1	M00026982C:D08
598	9/29/98	1494.001	71	RTA00001082F.h.17.1	M00027062C:C04
599	9/29/98	1494.001	72	RTA00001082F.p.14.1	M00027363D:A08
600	9/29/98	1494.001	73	RTA00001069F.j.04.1	M00027028A:B06
601	9/29/98	1494.001	74	RTA00001069F.p.21.1	M00027740C:C05
602	9/29/98	1494.001	75	RTA00001082F.e.07.1	M00026913D:G11
603	9/29/98	1494.001	76	RTA00001082F.d.23.3	M00026905A:G11
604	9/29/98	1494.001	77	RTA00001083F.b.18.1	M00027484A:G03
605	9/29/98	1494.001	78	RTA00001069F.o.06.1	M00027396A:F07
606	9/29/98	1494.001	79	RTA00001082F.p.01.1	M00027343B:H05
607	9/29/98	1494.001	80	RTA00001082F.p.11.1	M00027356A:H02
608	9/29/98	1494.001	81	RTA00001083F.f.19.1	M00027759B:E11

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
609	9/29/98	1494.001	82	RTA00001082F.i.04.1	M00027083D:F06
610	9/29/98	1494.001	83	RTA00001082F.p.12.1	M00027357D:A02
611	9/29/98	1494.001	84	RTA00001082F.d.15.3	M00026882A:E07
612	9/29/98	1494.001	85	RTA00001082F.i.20.1	M00027115B:G04
613	9/29/98	1494.001	86	RTA00001069F.d.03.1	M00023401C:D12
614	9/29/98	1494.001	87	RTA00001082F.e.10.1	M00026928A:B06
615	9/29/98	1494.001	88	RTA00001082F.a.07.1	M00023295B:C03
616	9/29/98	1494.001	89	RTA00001069F.n.15.1	M00027329A:H04
617	9/29/98	1494.001	90	RTA00001082F.d.08.3	M00026872A:C10
618	9/29/98	1494.001	91	RTA00001083F.f.13.1	M00027728A:B03
619	9/29/98	1494.001	92	RTA00001082F.b.03.1	M00023340B:H12
620	9/29/98	1494.001	93	RTA00001069F.b.09.1	M00023321B:F06
621	9/29/98	1494.001	94	RTA00001082F.l.20.1	M00027202B:B09
622	9/29/98	1494.001	95	RTA00001083F.c.14.1	M00027528A:G03
623	9/29/98	1494.001	96	RTA00001069F.c.07.1	M00023369D:C05
624	9/29/98	1494.001	97	RTA00001083F.d.16.1	M00027598C:D06
625	9/29/98	1494.001	98	RTA00001069F.e.22.1	M00026858C:H05
626	9/29/98	1494.001	99	RTA00001082F.j.10.1	M00027137C:A03
627	9/29/98	1494.001	100	RTA00001069F.b.01.1	M00023301B:C01
628	9/29/98	1494.001	101	RTA00001069F.j.20.1	M00027066A:A04
629	9/29/98	1494.001	102	RTA00001069F.e.24.1	M00026861A:B05
630	9/29/98	1494.001	103	RTA00001069F.b.08.1	M00023321A:F07
631	9/29/98	1494.001	104	RTA00001069F.k.16.1	M00027131A:H02
632	9/29/98	1494.001	105	RTA00001069F.j.22.1	M00027072C:A11
633	9/29/98	1494.001	106	RTA00001069F.j.07.1	M00027036B:D07
634	9/29/98	1494.001	107	RTA00001083F.c.20.1	M00027551C:B07
635	9/29/98	1494.001	108	RTA00001069F.l.11.1	M00027169D:H06
636	9/29/98	1494.001	109	RTA00001069F.c.03.1	M00023363C:A04
637	9/29/98	1494.001	110	RTA00001069F.l.14.1	M00027175D:A05
638	9/29/98	1494.001	111	RTA00001083F.c.10.1	M00027518B:B07
639	9/29/98	1494.001	112	RTA00001082F.a.04.1	M00023287A:D08
640	9/29/98	1494.001	113	RTA00001069F.m.13.1	M00027225B:D03
641	9/29/98	1494.001	114	RTA00001082F.n.08.1	M00027250A:C04
642	9/29/98	1494.001	115	RTA00001069F.e.09.1	M00026819B:E02
643	9/29/98	1494.001	116	RTA00001082F.p.18.1	M00027369A:B03
644	9/29/98	1494.001	117	RTA00001082F.d.24.3	M00026906B:G03
645	9/29/98	1494.001	118	RTA00001069F.c.23.1	M00023398D:F10
646	9/29/98	1494.001	119	RTA00001069F.b.19.1	M00023340B:B07

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
647	9/29/98	1494.001	120	RTA00001082F.n.03.1	M00027237C:D04
648	9/29/98	1494.001	121	RTA00001069F.a.13.1	M00023289D:E06
649	9/29/98	1494.001	122	RTA00001069F.e.16.1	M00026846C:B01
650	9/29/98	1494.001	123	RTA00001069F.p.04.1	M00027603C:E02
651	9/29/98	1494.001	124	RTA00001069F.m.21.1	M00027248D:D01
652	9/29/98	1494.001	125	RTA00001082F.h.14.1	M00027056B:H07
653	9/29/98	1494.001	126	RTA00001069F.p.03.1	M00027592D:C05
654	9/29/98	1494.001	127	RTA00001069F.n.02.1	M00027266C:G12
655	9/29/98	1494.001	128	RTA00001082F.m.01.1	M00027209D:B09
656	9/29/98	1494.001	129	RTA00001083F.e.09.1	M00027628D:D08
657	9/29/98	1494.001	130	RTA00001069F.d.18.1	M00023432D:F09
658	9/29/98	1494.001	131	RTA00001069F.e.06.1	M00026810A:H04
659	9/29/98	1494.001	132	RTA00001069F.e.05.1	M00026809C:D10
660	9/29/98	1494.001	133	RTA00001083F.c.05.1	M00027502C:H02
661	9/29/98	1494.001	134	RTA00001069F.c.10.1	M00023373A:D01
662	9/29/98	1494.001	135	RTA00001082F.k.10.1	M00027164A:A09
663	9/29/98	1494.001	136	RTA00001083F.c.07.1	M00027507C:C06
664	9/29/98	1494.001	137	RTA00001082F.j.15.1	M00027142A:C01
665	10/8/98	1495.001	1	RTA00001079F.j.08.1	M00022217B:E03
666	10/8/98	1495.001	2	RTA00001081F.h.04.1	M00022854D:C04
667	10/8/98	1495.001	3	RTA00001078F.h.08.1	M00021624B:D03
668	10/8/98	1495.001	4	RTA00001079F.b.12.1	M00022056C:D12
669	10/8/98	1495.001	5	RTA00001066F.o.03.1	M00022074A:F05
670	10/8/98	1495.001	6	RTA00001067F.p.05.1	M00022640B:G10
671	10/8/98	1495.001	7	RTA00001079F.l.05.1	M00022260C:H07
672	10/8/98	1495.001	8	RTA00001078F.f.17.1	M00008083A:H11
673	10/8/98	1495.001	9	RTA00001079F.l.04.1	M00022259A:D04
674	10/8/98	1495.001	10	RTA00001079F.m.19.1	M00022368C:C11
675	10/8/98	1495.001	11	RTA00001081F.f.08.1	M00022831C:F11
676	10/8/98	1495.001	12	RTA00001079F.e.13.1	M00022113B:A12
677	10/8/98	1495.001	13	RTA00001081F.f.21.1	M00022838B:E05
678	10/8/98	1495.001	14	RTA00001079F.g.11.1	M00022152A:G05
679	10/8/98	1495.001	15	RTA00001067F.i.05.1	M00022392C:H06
680	10/8/98	1495.001	16	RTA00001067F.n.01.1	M00022561B:B09
681	10/8/98	1495.001	17	RTA00001080F.i.20.1	M00022569D:H03
682	10/8/98	1495.001	18	RTA00001081F.p.04.1	M00023096A:F03
683	10/8/98	1495.001	19	RTA00001078F.d.04.1	M00008023A:B03
684	10/8/98	1495.001	20	RTA00001080F.h.09.1	M00022546B:F12

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
685	10/8/98	1495.001	21	RTA00000631F.a.10.3	M00022362D:G11
686	10/8/98	1495.001	22	RTA00001078F.f.15.1	M00008082B:H10
687	10/8/98	1495.001	23	RTA00001078F.a.11.1	M00007948D:F08
688	10/8/98	1495.001	24	RTA00001078F.e.08.1	M00008052C:G11
689	10/8/98	1495.001	25	RTA00001078F.c.08.1	M00008012D:E07
690	10/8/98	1495.001	26	RTA00001078F.b.18.1	M00008001B:E11
691	10/8/98	1495.001	27	RTA00001078F.d.08.1	M00008023C:A06
692	10/8/98	1495.001	28	RTA00001080F.p.19.1	M00022711B:A05
693	10/8/98	1495.001	29	RTA00001078F.a.17.1	M00007965C:B02
694	10/8/98	1495.001	30	RTA00001078F.n.22.2	M00021958A:A04
695	10/8/98	1495.001	31	RTA00001079F.d.12.1	M00022090D:B03
696	10/8/98	1495.001	32	RTA00001078F.j.16.1	M00021696C:E02
697	10/8/98	1495.001	33	RTA00001080F.n.06.1	M00022655A:F09
698	10/8/98	1495.001	34	RTA00001067F.d.16.1	M00022214A:D01
699	10/8/98	1495.001	35	RTA00001078F.l.03.2	M00021865B:F06
700	10/8/98	1495.001	36	RTA00001080F.o.02.1	M00022684B:F11
701	10/8/98	1495.001	37	RTA00001067F.p.15.1	M00022652B:G06
702	10/8/98	1495.001	38	RTA00001079F.d.16.1	M00022094A:A09
703	10/8/98	1495.001	39	RTA00001068F.c.17.1	M00022826A:C08
704	10/8/98	1495.001	40	RTA00001080F.g.05.1	M00022527D:A09
705	10/8/98	1495.001	41	RTA00001081F.e.07.1	M00022813C:B09
706	10/8/98	1495.001	42	RTA00001066F.g.16.1	M00021653C:B06
707	10/8/98	1495.001	43	RTA00001066F.l.05.1	M00021972A:C10
708	10/8/98	1495.001	44	RTA00001066F.h.16.1	M00021691B:E04
709	10/8/98	1495.001	45	RTA00001081F.g.13.1	M00022844C:A01
710	10/8/98	1495.001	46	RTA00001067F.p.07.1	M00022641C:H03
711	10/8/98	1495.001	47	RTA00001080F.g.02.1	M00022525C:E09
712	10/8/98	1495.001	48	RTA00001080F.i.02.1	M00022559D:F10
713	10/8/98	1495.001	49	RTA00001080F.g.22.1	M00022541D:G06
714	10/8/98	1495.001	50	RTA00001067F.d.20.1	M00022216C:H02
715	10/8/98	1495.001	51	RTA00001079F.k.17.1	M00022252A:C01
716	10/8/98	1495.001	52	RTA00001068F.d.04.1	M00022838A:H05
717	10/8/98	1495.001	53	RTA00001079F.n.11.1	M00022377A:E02
718	10/8/98	1495.001	54	RTA00001066F.d.22.1	M00008053D:E09
719	10/8/98	1495.001	55	RTA00001068F.f.08.1	M00023002A:C02
720	10/8/98	1495.001	56	RTA00001081F.o.16.1	M00023038D:D04
721	10/8/98	1495.001	57	RTA00001080F.f.18.1	M00022518C:C04
722	10/8/98	1495.001	58	RTA00001080F.a.16.1	M00022434D:B06

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
723	10/8/98	1495.001	59	RTA00001080F.j.18.1	M00022590D:E08
724	10/8/98	1495.001	60	RTA00001080F.n.11.1	M00022659B:C01
725	10/8/98	1495.001	61	RTA00001078F.e.01.1	M00008048C:A08
726	10/8/98	1495.001	62	RTA00001078F.b.07.1	M00007992A:G04
727	10/8/98	1495.001	63	RTA00001078F.b.01.1	M00007985C:G07
728	10/8/98	1495.001	64	RTA00001080F.n.14.1	M00022664A:E04
729	10/8/98	1495.001	65	RTA00001078F.o.21.2	M00021980A:F03
730	10/8/98	1495.001	66	RTA00001078F.c.06.1	M00008012B:C05
731	10/8/98	1495.001	67	RTA00001080F.o.15.1	M00022695D:B02
732	10/8/98	1495.001	68	RTA00001080F.o.16.1	M00022696A:H03
733	10/8/98	1495.001	69	RTA00001081F.a.07.2	M00022720A:C01
734	10/8/98	1495.001	70	RTA00001078F.f.22.1	M00008089C:B08
735	10/8/98	1495.001	71	RTA00001078F.g.02.1	M00008093C:G08
736	10/8/98	1495.001	72	RTA00001078F.j.13.2	M00021689A:G05
737	10/8/98	1495.001	73	RTA00001078F.l.02.2	M00021864C:C07
738	10/8/98	1495.001	74	RTA00001078F.i.14.2	M00021667C:G10
739	10/8/98	1495.001	75	RTA00001079F.d.04.1	M00022087A:D01
740	10/8/98	1495.001	76	RTA00001079F.l.09.1	M00022263A:C01
741	10/8/98	1495.001	77	RTA00001067F.o.19.1	M00022627B:D01
742	10/8/98	1495.001	78	RTA00001068F.b.01.1	M00022714B:D04
743	10/8/98	1495.001	79	RTA00001079F.f.07.1	M00022128A:C05
744	10/8/98	1495.001	80	RTA00001068F.a.03.1	M00022669D:G07
745	10/8/98	1495.001	81	RTA00001066F.f.03.1	M00008088D:B01
746	10/8/98	1495.001	82	RTA00001067F.o.18.1	M00022627A:A02
747	10/8/98	1495.001	83	RTA00001079F.k.12.1	M00022249C:G09
748	10/8/98	1495.001	84	RTA00001081F.g.07.1	M00022843A:D02
749	10/8/98	1495.001	85	RTA00001079F.j.01.1	M00022214A:H05
750	10/8/98	1495.001	86	RTA00001067F.p.10.1	M00022648D:G11
751	10/8/98	1495.001	87	RTA00001081F.f.16.1	M00022836C:A07
752	10/8/98	1495.001	88	RTA00001080F.i.05.1	M00022561D:E06
753	10/8/98	1495.001	89	RTA00001067F.l.02.1	M00022490B:G12
754	10/8/98	1495.001	90	RTA00001068F.a.23.1	M00022709A:G02
755	10/8/98	1495.001	91	RTA00001067F.d.18.1	M00022214C:E09
756	10/8/98	1495.001	92	RTA00001066F.o.05.1	M00022077D:A12
757	10/8/98	1495.001	93	RTA00001066F.m.08.1	M00022015D:C11
758	10/8/98	1495.001	94	RTA00001066F.b.12.1	M00007978B:C04
759	10/8/98	1495.001	95	RTA00001066F.c.08.1	M00008002B:F09
760	10/8/98	1495.001	96	RTA00001081F.p.05.1	M00023096C:A03

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
761	10/8/98	1495.001	97	RTA00001081F.c.01.1	M00022746D:D05
762	10/8/98	1495.001	98	RTA00001079F.m.23.1	M00022370A:G07
763	10/8/98	1495.001	99	RTA00001079F.m.09.1	M00022300A:A05
764	10/8/98	1495.001	100	RTA00001081F.c.21.1	M00022785C:B10
765	10/8/98	1495.001	101	RTA00001079F.o.04.1	M00022383C:F05
766	10/8/98	1495.001	102	RTA00001080F.b.10.1	M00022449D:B05
767	10/8/98	1495.001	103	RTA00001078F.c.09.1	M00008012D:H04
768	10/8/98	1495.001	104	RTA00001078F.d.19.1	M00008044C:A05
769	10/8/98	1495.001	105	RTA00001081F.a.11.2	M00022722D:C07
770	10/8/98	1495.001	106	RTA00001080F.n.15.1	M00022664C:G10
771	10/8/98	1495.001	107	RTA00001078F.a.09.1	M00007941D:D07
772	10/8/98	1495.001	108	RTA00001078F.g.20.1	M00021614A:C09
773	10/8/98	1495.001	109	RTA00001066F.h.23.1	M00021841A:E11
774	10/8/98	1495.001	110	RTA00001081F.l.11.2	M00022922D:G06
775	10/8/98	1495.001	111	RTA00001079F.d.18.1	M00022096B:D10
776	10/8/98	1495.001	112	RTA00001066F.f.21.1	M00008100D:C08
777	10/8/98	1495.001	113	RTA00001078F.j.06.1	M00021680D:H08
778	10/8/98	1495.001	114	RTA00001067F.d.08.1	M00022205A:C02
779	10/8/98	1495.001	115	RTA00001068F.b.05.1	M00022717C:F05
780	10/8/98	1495.001	116	RTA00001079F.c.05.1	M00022071D:C08
781	10/8/98	1495.001	117	RTA00001078F.k.10.2	M00021852C:D12
782	10/8/98	1495.001	118	RTA00001081F.i.18.2	M00022884D:A07
783	10/8/98	1495.001	119	RTA00001066F.b.21.1	M00007996C:B11
784	10/8/98	1495.001	120	RTA00001066F.i.08.1	M00021851D:H06
785	10/8/98	1495.001	121	RTA00001068F.e.08.1	M00022915C:C09
786	10/8/98	1495.001	122	RTA00001079F.j.15.1	M00022220B:B06
787	10/8/98	1495.001	123	RTA00001078F.j.18.2	M00021698A:H03
788	10/8/98	1495.001	124	RTA00001066F.b.09.1	M00007977B:C11
789	10/8/98	1495.001	125	RTA00001079F.i.20.1	M00022207C:C01
790	10/8/98	1495.001	126	RTA00001080F.e.15.1	M00022506D:B03
791	10/8/98	1495.001	127	RTA00001080F.l.03.1	M00022617B:A01
792	10/8/98	1495.001	128	RTA00001080F.e.10.1	M00022501D:A09
793	10/8/98	1495.001	129	RTA00001067F.c.22.1	M00022184D:F07
794	10/8/98	1495.001	130	RTA00001081F.p.11.1	M00023097A:C03
795	10/8/98	1495.001	131	RTA00001081F.p.08.1	M00023096D:B11
796	10/8/98	1495.001	132	RTA00001080F.c.19.1	M00022471D:A05
797	10/8/98	1495.001	133	RTA00001081F.b.06.1	M00022736B:B03
798	10/8/98	1495.001	134	RTA00001081F.m.22.1	M00022983A:H04

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
799	10/8/98	1495.001	135	RTA00001081F.d.11.1	M00022801A:G04
800	10/8/98	1495.001	136	RTA00001081F.n.13.1	M00023002D:C12
801	10/8/98	1495.001	137	RTA00001067F.d.17.1	M00022214C:C11
802	10/8/98	1495.001	138	RTA00001081F.c.13.1	M00022772A:A06
803	10/8/98	1495.001	139	RTA00001078F.b.19.1	M00008001D:F11
804	10/8/98	1495.001	140	RTA00001078F.a.04.1	M00007931A:B07
805	10/8/98	1495.001	141	RTA00001078F.b.16.1	M00008000D:G11
806	10/8/98	1495.001	142	RTA00001078F.b.04.1	M00007987A:D10
807	10/8/98	1495.001	143	RTA00001078F.d.18.1	M00008044B:F07
808	10/8/98	1495.001	144	RTA00001068F.e.05.1	M00022904D:D04
809	10/8/98	1495.001	145	RTA00001078F.i.18.1	M00021674A:B07
810	10/8/98	1495.001	146	RTA00001066F.e.01.1	M00008054C:C03
811	10/8/98	1495.001	147	RTA00001078F.n.14.2	M00021949D:A05
812	10/8/98	1495.001	148	RTA00001067F.i.17.1	M00022413B:D07
813	10/8/98	1495.001	149	RTA00001079F.l.19.1	M00022278C:E04
814	10/8/98	1495.001	150	RTA00001081F.l.12.2	M00022923A:A09
815	10/8/98	1495.001	151	RTA00001067F.j.03.1	M00022420B:C08
816	10/8/98	1495.001	152	RTA00001068F.d.19.1	M00022898C:H07
817	10/8/98	1495.001	153	RTA00001081F.g.23.1	M00022853D:C05
818	10/8/98	1495.001	154	RTA00001081F.h.16.1	M00022860A:A07
819	10/8/98	1495.001	155	RTA00001079F.i.05.1	M00022192B:H07
820	10/8/98	1495.001	156	RTA00001068F.f.12.1	M00023012A:C06
821	10/8/98	1495.001	157	RTA00001067F.e.09.1	M00022235D:F07
822	10/8/98	1495.001	158	RTA00001066F.m.10.1	M00022018B:E09
823	10/8/98	1495.001	159	RTA00001080F.j.19.1	M00022591C:F03
824	10/8/98	1495.001	160	RTA00001080F.f.07.1	M00022513C:G04
825	10/8/98	1495.001	161	RTA00001080F.e.09.1	M00022500B:D01
826	10/8/98	1495.001	162	RTA00001080F.e.19.1	M00022509D:A12
827	10/8/98	1495.001	163	RTA00001066F.a.13.1	M00007948B:B07
828	10/8/98	1495.001	164	RTA00001079F.p.14.1	M00022407D:G07
829	10/8/98	1495.001	165	RTA00001079F.p.03.1	M00022399C:B02
830	10/8/98	1495.001	166	RTA00001079F.n.22.1	M00022381B:C12
831	10/8/98	1495.001	167	RTA00001078F.a.06.1	M00007937C:E08
832	10/8/98	1495.001	168	RTA00001078F.a.19.1	M00007973D:B03
833	10/8/98	1495.001	169	RTA00001078F.b.15.1	M00008000D:B06
834	10/8/98	1495.001	170	RTA00001079F.c.15.1	M00022078B:B04
835	10/8/98	1495.001	171	RTA00001079F.d.06.1	M00022088B:E05
836	10/8/98	1495.001	172	RTA00001067F.a.05.1	M00022118A:D08

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
837	10/8/98	1495.001	173	RTA00001078F.i.15.2	M00021668D:G09
838	10/8/98	1495.001	174	RTA00001066F.a.11.1	M00007947B:F07
839	10/8/98	1495.001	175	RTA00001078F.k.02.2	M00021846B:F05
840	10/8/98	1495.001	176	RTA00001066F.h.04.1	M00021669B:G02
841	10/8/98	1495.001	177	RTA00001066F.c.21.1	M00008015B:D08
842	10/8/98	1495.001	178	RTA00001080F.h.06.1	M00022544C:D08
843	10/8/98	1495.001	179	RTA00001067F.c.16.1	M00022177D:G02
844	10/8/98	1495.001	180	RTA00001080F.f.21.1	M00022522B:A05
845	10/8/98	1495.001	181	RTA00001080F.a.10.1	M00022425A:F11
846	10/8/98	1495.001	182	RTA00001081F.o.10.1	M00023034B:B10
847	10/8/98	1495.001	183	RTA00001078F.b.17.1	M00008001A:G11
848	10/8/98	1495.001	184	RTA00001078F.g.04.1	M00008094D:C02
849	10/8/98	1495.001	185	RTA00001080F.p.05.1	M00022704A:H08
850	10/8/98	1495.001	186	RTA00001067F.f.04.1	M00022256D:G11
851	10/8/98	1495.001	187	RTA00001066F.c.11.1	M00008003B:F09
852	10/8/98	1495.001	188	RTA00001081F.b.19.1	M00022743C:G05
853	10/8/98	1495.001	189	RTA00001081F.p.14.1	M00023097C:D10
854	10/8/98	1495.001	190	RTA00001067F.k.16.1	M00022467C:H07
855	10/8/98	1495.001	191	RTA00001081F.b.11.1	M00022737D:B02
856	10/8/98	1495.001	192	RTA00001080F.k.12.1	M00022601A:A09
857	10/8/98	1495.001	193	RTA00001066F.a.08.1	M00007943C:B02
858	10/8/98	1495.001	194	RTA00001081F.b.10.1	M00022737B:F12
859	10/8/98	1495.001	195	RTA00001080F.d.15.1	M00022488C:H02
860	10/8/98	1495.001	196	RTA00001079F.p.04.1	M00022399D:A07
861	10/8/98	1495.001	197	RTA00001067F.e.23.1	M00022251A:F07
862	10/8/98	1495.001	198	RTA00001068F.a.08.1	M00022684C:C12
863	10/8/98	1495.001	199	RTA00001078F.h.16.1	M00021628C:B09
864	10/8/98	1495.001	200	RTA00001081F.g.18.1	M00022848D:H09
865	10/8/98	1495.001	201	RTA00001081F.m.15.1	M00022968D:G06
866	10/8/98	1495.001	202	RTA00001067F.k.09.1	M00022459C:G05
867	10/8/98	1495.001	203	RTA00001080F.g.04.1	M00022527B:H05
868	10/8/98	1495.001	204	RTA00001081F.j.19.2	M00022902C:F11
869	10/8/98	1495.001	205	RTA00001081F.o.03.1	M00023023B:A05
870	10/8/98	1495.001	206	RTA00001079F.b.23.1	M00022067A:B03
871	10/8/98	1495.001	207	RTA00001078F.n.16.2	M00021951B:A01
872	10/8/98	1495.001	208	RTA00001067F.b.01.1	M00022134D:D12
873	10/8/98	1495.001	209	RTA00001080F.a.17.1	M00022435C:C05
874	10/8/98	1495.001	210	RTA00001080F.c.17.1	M00022469A:A05

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
875	10/8/98	1495.001	211	RTA00001068F.f.10.1	M00023003C:C10
876	10/8/98	1495.001	212	RTA00001081F.h.18.1	M00022861C:B04
877	10/8/98	1495.001	213	RTA00001066F.p.19.1	M00022106D:B06
878	10/8/98	1495.001	214	RTA00001080F.c.09.1	M00022464D:F12
879	10/8/98	1495.001	215	RTA00001078F.c.12.1	M00008014C:H01
880	10/8/98	1495.001	216	RTA00001080F.l.10.1	M00022622A:E08
881	10/8/98	1495.001	217	RTA00001078F.g.11.1	M00008099A:C12
882	10/8/98	1495.001	218	RTA00001068F.f.09.1	M00023003A:H01
883	10/8/98	1495.001	219	RTA00001067F.f.10.1	M00022261C:D06
884	10/8/98	1495.001	220	RTA00001080F.o.05.1	M00022687C:C11
885	10/8/98	1495.001	221	RTA00001078F.h.04.1	M00021620D:B06
886	10/8/98	1495.001	222	RTA00001078F.p.03.2	M00021981D:A11
887	10/8/98	1495.001	223	RTA00001080F.e.20.1	M00022510A:B09
888	10/8/98	1495.001	224	RTA00001078F.k.19.2	M00021861C:B08
889	10/8/98	1495.001	225	RTA00001078F.d.20.1	M00008045A:B05
890	10/8/98	1495.001	226	RTA00001078F.b.22.1	M00008006A:H02
891	10/8/98	1495.001	227	RTA00001068F.a.13.1	M00022701C:A05
892	10/8/98	1495.001	228	RTA00001080F.m.16.1	M00022641D:F08
893	10/8/98	1495.001	229	RTA00001080F.o.22.1	M00022702A:D10
894	10/8/98	1495.001	230	RTA00001080F.k.16.1	M00022604A:F06
895	10/8/98	1495.001	231	RTA00001067F.d.04.1	M00022199A:F09
896	10/8/98	1495.001	232	RTA00001067F.k.10.1	M00022460C:E12
897	10/8/98	1495.001	233	RTA00001078F.n.04.2	M00021931B:F04
898	10/8/98	1495.001	234	RTA00001078F.n.07.2	M00021945A:B04
899	10/8/98	1495.001	235	RTA00001081F.a.16.1	M00022725D:G05
900	10/8/98	1495.001	236	RTA00001078F.l.13.2	M00021879B:C11
901	10/8/98	1495.001	237	RTA00001078F.f.13.1	M00008082B:C05
902	10/8/98	1495.001	238	RTA00001079F.d.05.1	M00022087D:F12
903	10/8/98	1495.001	239	RTA00001067F.i.13.1	M00022406C:G03
904	10/8/98	1495.001	240	RTA00001068F.d.23.1	M00022902B:F10
905	10/8/98	1495.001	241	RTA00001078F.c.13.1	M00008014D:A11
906	10/8/98	1495.001	242	RTA00001078F.a.18.1	M00007969B:E10
907	10/8/98	1495.001	243	RTA00001068F.b.23.1	M00022765B:E03
908	10/8/98	1495.001	244	RTA00001078F.f.21.1	M00008085B:G01
909	10/8/98	1495.001	245	RTA00001067F.b.15.1	M00022144D:D09
910	10/8/98	1495.001	246	RTA00001078F.o.04.2	M00021963C:H04
911	10/8/98	1495.001	247	RTA00001081F.e.14.1	M00022817D:B09
912	10/8/98	1495.001	248	RTA00001078F.k.04.2	M00021847B:A09

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
913	10/8/98	1495.001	249	RTA00001079F.g.15.2	M00022158C:C08
914	10/8/98	1495.001	250	RTA00001067F.k.23.1	M00022477C:C07
915	10/8/98	1495.001	251	RTA00001079F.h.08.2	M00022176A:F02
916	10/8/98	1495.001	252	RTA00001078F.d.17.1	M00008028D:B01
917	10/8/98	1495.001	253	RTA00001067F.d.07.1	M00022203B:A05
918	10/8/98	1495.001	254	RTA00001068F.e.04.1	M00022903D:H02
919	10/8/98	1495.001	255	RTA00001068F.a.06.1	M00022682A:F10
920	10/8/98	1495.001	256	RTA00001078F.e.10.1	M00008054C:E07
921	10/8/98	1495.001	257	RTA00001079F.b.11.1	M00022056B:G12
922	10/8/98	1495.001	258	RTA00001066F.h.11.1	M00021676B:B12
923	10/8/98	1495.001	259	RTA00001079F.d.01.1	M00022084B:C03
924	10/8/98	1495.001	260	RTA00001067F.g.14.1	M00022363C:D03
925	10/8/98	1495.001	261	RTA00001066F.g.06.1	M00021625B:G07
926	10/8/98	1495.001	262	RTA00001081F.j.09.2	M00022893D:C06
927	10/8/98	1495.001	263	RTA00001068F.e.19.1	M00022963A:E07
928	10/8/98	1495.001	264	RTA00001079F.l.21.1	M00022282A:A11
929	10/8/98	1495.001	265	RTA00001078F.h.09.1	M00021624B:E11
930	10/8/98	1495.001	266	RTA00001078F.d.16.1	M00008027D:H09
931	10/8/98	1495.001	267	RTA00001079F.g.22.2	M00022167B:H02
932	10/8/98	1495.001	268	RTA00001066F.e.15.1	M00008075D:B01
933	10/8/98	1495.001	269	RTA00001080F.g.16.1	M00022538D:B02
934	10/8/98	1495.001	270	RTA00001080F.b.07.1	M00022447A:H06
935	10/8/98	1495.001	271	RTA00001078F.n.21.2	M00021958A:A03
936	10/8/98	1495.001	272	RTA00001078F.b.12.1	M00007998C:B04
937	10/8/98	1495.001	273	RTA00001066F.p.01.2	M00022099C:A10
938	10/8/98	1495.001	274	RTA00001066F.o.22.1	M00022095C:F03
939	10/8/98	1495.001	275	RTA00001080F.i.19.1	M00022568B:D03
940	10/8/98	1495.001	276	RTA00001079F.g.01.1	M00022138C:B07
941	10/8/98	1495.001	277	RTA00001079F.e.02.1	M00022102D:A10
942	10/8/98	1495.001	278	RTA00001079F.k.01.1	M00022233C:D11
943	10/8/98	1495.001	279	RTA00001079F.o.11.1	M00022386D:C04
944	10/8/98	1495.001	280	RTA00001068F.d.02.1	M00022834A:H02
945	10/8/98	1495.001	281	RTA00001078F.a.07.1	M00007939A:F06
946	10/8/98	1495.001	282	RTA00001081F.b.20.1	M00022743C:G06
947	10/8/98	1495.001	283	RTA00001067F.f.20.1	M00022273A:B03
948	10/8/98	1495.001	284	RTA00001079F.c.06.1	M00022072D:E12
949	10/8/98	1495.001	285	RTA00001068F.b.24.1	M00022768A:A10
950	10/8/98	1495.001	286	RTA00001080F.o.08.1	M00022691A:G01

Table 1A

	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
951	10/8/98	1495.001	287	RTA00001078F.j.10.2	M00021687C:A04
952	10/8/98	1495.001	288	RTA00001080F.b.03.1	M00022444B:C04
953	10/8/98	1495.001	289	RTA00001067F.e.13.1	M00022240C:B03
954	10/8/98	1495.001	290	RTA00001081F.h.05.1	M00022856A:B09
955	10/8/98	1495.001	291	RTA00001067F.f.01.1	M00022252C:A04
956	10/8/98	1495.001	292	RTA00001080F.g.23.1	M00022542A:B06
957	10/8/98	1495.001	293	RTA00001080F.h.16.1	M00022548A:F02
958	10/8/98	1495.001	294	RTA00001080F.f.15.1	M00022517C:B01
959	10/8/98	1495.001	295	RTA00001080F.f.06.1	M00022513C:E10
960	10/8/98	1495.001	296	RTA00001081F.a.04.2	M00022716A:C01
961	10/8/98	1495.001	297	RTA00001078F.p.16.2	M00022001B:H10
962	10/8/98	1495.001	298	RTA00001081F.b.03.1	M00022734C:A03
963	10/8/98	1495.001	299	RTA00001080F.a.21.1	M00022441B:A06
964	10/8/98	1495.001	300	RTA00001079F.f.05.1	M00022127C:E01
965	10/8/98	1495.001	301	RTA00001080F.n.23.1	M00022681D:H10
966	10/8/98	1495.001	302	RTA00001078F.c.18.1	M00008016C:E06
967	10/8/98	1495.001	303	RTA00001068F.a.11.1	M00022697A:C08
968	10/8/98	1495.001	304	RTA00001068F.g.09.1	M00023095C:A09
969	10/8/98	1495.001	305	RTA00001068F.a.22.1	M00022709A:C01
970	10/8/98	1495.001	306	RTA00001079F.h.09.2	M00022176D:F05
971	10/8/98	1495.001	307	RTA00001079F.h.01.2	M00022169A:E11
972	10/8/98	1495.001	308	RTA00001078F.g.07.1	M00008097C:E04
973	10/8/98	1495.001	309	RTA00001078F.m.08.2	M00021908B:F03
974	10/8/98	1495.001	310	RTA00001080F.a.03.1	M00022417B:C01
975	10/8/98	1495.001	311	RTA00001079F.o.06.1	M00022384B:E06
976	10/8/98	1495.001	312	RTA00001079F.p.06.1	M00022401C:G07
977	10/8/98	1495.001	313	RTA00001078F.p.18.2	M00022001D:E06
978	10/8/98	1495.001	314	RTA00001068F.a.17.1	M00022705B:F08
979	10/8/98	1495.001	315	RTA00001078F.a.10.1	M00007948C:G01
980	10/8/98	1495.001	316	RTA00001079F.h.20.2	M00022184D:H07
981	10/8/98	1495.001	317	RTA00001081F.n.03.1	M00022986B:C02
982	10/8/98	1495.001	318	RTA00001080F.c.04.1	M00022460D:C07

Table 1B

SEQ ID NO:	Sample Name	Clone ID
983	270.F5.sp6:145120	M00001401B:A02
984	344.C4.sp6:146251	M00023363C:A04
985	628.D9.sp6:157832	M00008028D:B01
986	628.F7.sp6:157854	M00008023C:A06
987	636.G12.sp6:158255	M00022077D:A12
988	653.F3.sp6:159004	M00023284B:G06
989	654.H6.sp6:159223	M00023369D:C05
990	655.B2.sp6:156468	M00023413D:F04
991	656.B11.sp6:159348	M00026905A:G11
992	661.C10.sp6:159743	M00027169D:H06
993	953.B04.sp6:185140	M00005434D:H02
994	270.F5.sp6:145120	M00001401B:A02
995	344.C4.sp6:146251	M00023363C:A04
996	655.B2.sp6:156468	M00023413D:F04

Table 1C

SEQ ID NO:	Sequence Name	THC Accession No.
997	RTA00001071F.i.23.3	AA173046
998	RTA00001079F.m.19.1	THC220786
999	RTA00001067F.i.05.1	THC233199
1000	RTA00001082F.o.01.1	THC178783
1001	RTA00001067F.n.01.1	AA173079
1002	RTA00001076F.b.13.1	AA554659
1003	RTA00001064F.p.03.1	AA432284
1004	RTA00001072F.g.05.2	H20612
1005	RTA00001064F.c.01.1	EST55879
1006	RTA00001083F.b.09.1	W30744
1007	RTA00001083F.c.03.1	THC205070
1008	RTA00001066F.h.16.1	EST14169
1009	RTA00001076F.n.10.1	THC144372
1010	RTA00001061F.e.17.1	N48670
1011	RTA00001071F.m.09.3	R56510
1012	RTA00001080F.g.02.1	THC77700
1013	RTA00001073F.i.02.2	Z46186
1014	RTA00001076F.j.14.1	THC144372
1015	RTA00001068F.d.04.1	AA011604
1016	RTA00001069F.o.11.1	AA576259
1017	RTA00001073F.k.01.1	R52934
1018	RTA00001080F.f.18.1	THC126698
1019	RTA00001075F.e.18.1	THC209874
1020	RTA00001076F.d.13.1	AA158197
1021	RTA00001065F.f.06.1	THC219476

SEQ ID NO:	Sequence Name	THC Accession No.
1022	RTA00001068F.b.01.1	THC151511
1023	RTA00001068F.a.03.1	THC220020
1024	RTA00001072F.b.09.2	AA554360
1025	RTA00001076F.i.09.1	EST20991
1026	RTA00001073F.l.04.1	AA527712
1027	RTA00001067F.d.18.1	THC198501
1028	RTA00001082F.b.03.1	THC218291
1029	RTA00001082F.l.20.1	THC204015
1030	RTA00001081F.c.21.1	THC203534
1031	RTA00001069F.b.08.1	THC234347
1032	RTA00001074F.f.09.1	N53623
1033	RTA00001066F.h.23.1	THC129284
1034	RTA00001064F.h.07.1	THC161794
1035	RTA00001066F.f.21.1	T92493
1036	RTA00001069F.m.13.1	AA148143
1037	RTA00001064F.d.14.1	THC138642
1038	RTA00001068F.e.08.1	AA633643
1039	RTA00001065F.d.19.1	THC227618
1040	RTA00001069F.e.06.1	T19066
1041	RTA00001069F.e.05.1	T19066
1042	RTA00001082F.j.15.1	THC226714
1043	RTA00001067F.i.17.1	EST83778
1044	RTA00001081F.l.12.2	AA121009
1045	RTA00001080F.e.19.1	T99190
1046	RTA00001065F.d.18.2	H59526
1047	RTA00001078F.a.06.1	AA453802
1048	RTA00001065F.a.21.1	THC86626
1049	RTA00001075F.a.02.1	AA632565
1050	RTA00001066F.c.21.1	AA465322
1051	RTA00001080F.h.06.1	THC232157
1052	RTA00001067F.b.01.1	EST79811
1053	RTA00001071F.l.19.1	THC208816
1054	RTA00001062F.f.01.1	THC105335
1055	RTA00001063F.g.18.1	THC205088
1056	RTA00001062F.j.18.1	THC220715
1057	RTA00001078F.b.22.1	THC232576
1058	RTA00001064F.a.09.2	THC171312
1059	RTA00001064F.k.20.2	THC200994
1060	RTA00001080F.m.16.1	EST62430
1061	RTA00001078F.n.04.2	THC231131
1062	RTA00001071F.p.07.1	AA524115
1063	RTA00001074F.k.15.1	AA053768
1064	RTA00001073F.g.22.1	THC146930

SEQ ID NO:	Sequence Name	THC Accession No.
1065	RTA00001067F.k.23.1	THC211481
1066	RTA00001068F.a.06.1	THC232664
1067	RTA00001067F.g.14.1	THC110314
1068	RTA00001072F.i.19.3	EST84170
1069	RTA00001079F.g.22.2	THC146930
1070	RTA00001061F.j.03.1	THC195525
1071	RTA00001072F.c.16.2	AA159011
1072	RTA00001061F.c.12.1	THC196151
1073	RTA00001072F.j.23.2	N99474
1074	RTA00001080F.f.06.1	R06925
1075	RTA00001080F.a.21.1	THC173393
1076	RTA00001068F.a.11.1	THC202663
1077	RTA00001078F.g.07.1	EST89489
1078	RTA00001078F.m.08.2	THC233725
1079	RTA00001068F.a.17.1	N86176

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1	NM_005757.1	Homo sapiens C3H-type zinc finger protein; similar to D. melanogaster muscleblind B protein (MBLL) mRNA >gi|3779239|gb|AF061261|AF061261 Homo sapiens zinc finger protein (MBLL) mRNA, complete cds	7.00E-99
5	M86697	Peptostreptococcus magnus protein L gene, complete cds.	1.90E+00
6	AF080255.1	Homo sapiens lodestar protein mRNA, complete cds	1.00E-37
7	Z95310	Caenorhabditis elegans cosmid H40L08, complete sequence [Caenorhabditis elegans]	2.00E+00
9	AF124981.1	Bombyx mori nuclear receptor GRF (GRF) mRNA, complete cds	1.90E+00
10	U43663	Xenopus laevis transposon TXr.11 transposase pseudogene, complete cds	4.20E+00
11	AE001495	Helicobacter pylori, strain J99 section 56 of 132 of the complete genome	2.00E+00
12	AB031040.1	Mus musculus mLhx6.1a mRNA for LIM-homeodomain (LHX) protein 6.1a, complete cds	1.00E-79
13	AF132973.1	Homo sapiens CGI-39 protein mRNA, complete cds	2.00E-30
14	L81907	Homo sapiens (subclone 1_c12 from P1 H69) DNA sequence	2.00E+00
15	AE001543	Helicobacter pylori, strain J99 section 104 of 132 of the complete genome	8.00E-03
16	L42167	Mus musculus (clone R24) rds gene, partial cds	4.70E-01
17	U58870	Bos taurus carbonic anhydrase IV mRNA, complete cds	6.80E-01
18	AB025187.1	Oryza sativa mRNA for cytochrome c oxidase subunit 6b-1, complete cds	2.30E-01
19	AE000723	Aquifex aeolicus section 55 of 109 of the complete genome	6.80E-01
20	U72058	Mus musculus chloride channel regulator (Icln) gene, exon 2 and partial cds	6.80E-01
21	U24698	Aspergillus parasiticus norsolorinic acid reductase (nor) gene, complete cds	6.50E-01
22	AB014528	Homo sapiens mRNA for KIAA0628 protein, complete cds	0.00E+00
23	X90691	M.morganii DNA for orf3, orf4, orf5, orf6, orf7, orf8, orf9, and rumA & rumB genes	2.00E+00
25	U24098	Macaca fascicularis eosinophil cationic protein gene, complete cds	6.60E-01
27	U19355	Rattus norvegicus satellite sequence d0Mco3.	6.60E-01
28	U39655	Caenorhabditis elegans cosmid C46F4	1.90E+00
29	M34463	Rat S-adenosylmethionine decarboxylase (AMDP1) pseudogene, complete cds.	1.90E+00

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
30	AB028898.1	Mus musculus mRNA for U8, complete cds	7.00E-43
31	M19547	D.melanogaster (strain Af-S) alcohol dehydrogenase gene (allele Adh-S), complete cds.	2.00E+00
32	AB029001.1	Homo sapiens mRNA for KIAA1078 protein, partial cds	e-158
33	AF063668.1	Mus musculus type XIII collagen (col13a1) gene, exon 3	2.10E+00
34	X07356	Chicken nicotinic acetylcholine receptor non-alpha gene exon 5	6.60E-01
35	U69609	Human transcriptional repressor (GCF2) mRNA, complete cds	1.80E+00
36	AF041861	Mus musculus synaptojanin 2 isoform zeta mRNA, partial cds	1.90E+00
37	AB028975.1	Homo sapiens mRNA for KIAA1052 protein, complete cds	1.90E+00
38	AF046000	Mus musculus rod cGMP phosphodiesterase delta subunit (Pde6d) gene, complete cds	5.50E-01
39	L09705	Human DNA sequence.	6.10E-01
41	AF052692	Homo sapiens connexin 31 (GJB3) mRNA, complete cds	e-132
42	Z80214	Caenorhabditis elegans cosmid C27D8, complete sequence [Caenorhabditis elegans]	4.70E-01
43	M95520	Streptococcus canis (group G) albumin-binding protein gene, partial cds.	2.30E-01
44	AE001392	Plasmodium falciparum chromosome 2, section 29 of 73 of the complete sequence	7.70E-02
45	AF112187	Mus musculus epithelial sodium channel gamma subunit mRNA, complete cds	2.10E+00
46	M31616	O.sativa ADPglucose pyrophosphorylase gene, complete cds.	2.30E-01
47	U20734	Human transcription factor junB (junB) gene, 5' region and complete cds.	2.30E-01
48	U35782	Anopheles bwambae 12S ribosomal RNA, D-loop, and tRNA-Ile mitochondrial genes, partial sequence.	2.30E-01
49	AF138280.1	Gallus gallus chondromodulin-I mRNA, complete cds	3.00E-03
50	AE001392	Plasmodium falciparum chromosome 2, section 29 of 73 of the complete sequence	7.60E-02
52	D45385	Pokeweed mRNA for polyphenol oxidase, complete cds	2.20E-01
53	J04804	C.elegans vinculin (deb-1) gene, complete cds.	2.20E-01
54	M34431	Human PVT-IGLC fusion protein mRNA, 5' end.	6.70E-01
55	L05634	Bacillus subtilis ORF1, 3' end; wall-associated protein (walA) gene, complete cds; complete ORF3.	6.50E-01

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
56	X82430	E.coli transposable element IS1294	2.20E-01
57	AJ224356	Solanum lycopersicon tDET1 gene	6.70E-01
58	U07792	Human tyrosine kinase (TXK) gene, exon 8, partial cds. >gi|1161352|gb|U34371|HSTECTXT05 Human tyrosine kinase TXK (txk) gene, exon 5.	6.60E-01
59	D87559	Bos taurus mRNA for platelet-activating factor acetylhydrolase 2, complete cds	2.20E-01
60	U45427	Borrelia burgdorferi 2.9-7 locus, ORF-A-D, REV, and lipoprotein (LPA and LPB) genes, complete cds	6.60E-01
61	AB018343.1	Homo sapiens mRNA for KIAA0800 protein, complete cds	e-177
62	X92186	M.musculus 11beta-hydroxysteroid dehydrogenase type 1 gene	0.66
63	U51899	Human kappa-casein gene, complete cds	0.65
64	S62069	cathepsin B {5' region} [human, Genomic, 886 nt, segment 1 of 2]	0.22
65	M26198	Bovine ASS mRNA encoding argininosuccinate synthetase, complete cds.	0.22
66	AE001628	Chlamydia pneumoniae section 44 of 103 of the complete genome	2.20E-01
67	AF097906	Rana catesbeiana myosin heavy chain (MHC-3) mRNA, partial cds	6.40E-01
68	AE001023	Archaeoglobus fulgidus section 84 of 172 of the complete genome	2.30E-01
69	U11682	Trypanoplasma borelli mitochondrion cytochrome oxidase subunit 1 (cox1), cytochrome oxidase subunit 2 and complete 9S rRNA gene and partial 12S rRNA gene.	6.40E-01
70	J03267	Rat atrial natriuretic factor (ANF) gene, 5' end.	6.40E-01
71	AL034546.5	Human DNA sequence from clone 898I4 on chromosome 22q13.33. Contains a GSS and a putative CpG island, complete sequence [Homo sapiens]	6.20E-01
72	U78730	Homo sapiens mad protein homolog Smad2 gene, exon 7	1.90E-01
73	D87686.1	Homo sapiens mRNA for KIAA0017 protein, complete cds	e-175
74	AF085361.1	Homo sapiens HSPC032 mRNA, complete cds	2.00E-55
75	AF168786.1	Sorghum bicolor soluble starch synthase mRNA, partial cds	2.50E-02
76	Z99102	Caenorhabditis elegans cosmid B0331, complete sequence [Caenorhabditis elegans]	7.40E-02
77	AE001247	Treponema pallidum section 63 of 87 of the complete genome	2.30E-01

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
78	U20734	Human transcription factor junB (junB) gene, 5' region and complete cds.	5.00E-08
79	X92112	G.gallus mRNA for guanylate-binding protein	7.50E-02
80	AF043692	Caenorhabditis elegans cosmid C17F3	2.00E+00
81	D88260	Pisum sativum PsCHS4 gene for chalcone synthase, complete cds	6.70E-01
82	D87433	Human mRNA for KIAA0246 gene, partial cds	2.30E-01
83	X70301	S.lemnae internal telomeric sequence maa81	2.30E-01
84	AB018249	Homo sapiens gene for CC chemokine LEC, complete cds	7.50E-02
85	D32072	Mouse mRNA for an isoform of TGF-b type II receptor	7.40E-02
86	AB018317.1	Homo sapiens mRNA for KIAA0774 protein, partial cds	1.90E+00
87	Z46372	R.norvegicus RNA for DNA topoisomerase II	7.20E-02
88	NM_003958.1	Homo sapiens C3HC4-type zinc finger protein sapiens mRNA for KIAA0646 protein, complete cds	6.50E-01
89	AF005655	Eschscholzia californica berberine bridge enzyme (bbe1) gene, complete cds	7.70E-02
90	AF042192	Xenopus laevis paraxial protocadherin mRNA, complete cds	6.20E-01
91	Y12002	N.crassa DNA for protein kinase C homologue	2.20E-01
92	AF077697	HIV-1 isolate DW.s.0 from Switzerland pol protein (pol) gene, partial cds	2.00E-01
93	L31848	Homo sapiens serine/threonine kinase receptor 2	6.00E-11
94	AF047707	Rattus norvegicus UDP-glucose:ceramide glycosyltransferase mRNA, complete cds	6.00E-01
95	X92112	G.gallus mRNA for guanylate-binding protein	7.10E-02
96	X82333	H.sapiens IRLB gene (exon1-3)	5.30E-02
97	AJ228139.2	Homo sapiens mRNA for LETKI precursor	2.00E-97
98	M13075	Human albumin gene, exon 1 and 5' flank.	1.40E+00
99	AF025430	Papaver somniferum berberine bridge enzyme	2.90E-01
100	X92346	M.musculus mRNA for CART1 protein	1.70E-02
102	AE001367	Plasmodium falciparum chromosome 2, section 4 of 73 of the complete sequence	2.50E-02
103	AB014524	Homo sapiens mRNA for KIAA0624 protein, partial cds	0.00E+00
104	AB007546	Homo sapiens gene for LECT2, complete cds	2.20E-01
105	AF060492	Buchnera aphidicola succinyl-diaminopimelate aminotransferase (dapD) gene, partial cds; periplasmic serine protease (htrA), hypothetical protein, acetohydroxy acid synthase large subunit (ilvI), acetohydroxy acid synthas...	7.50E-02
106	D16360	Human DNA for plasma glutathione peroxidase, exon 1	2.50E-02

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
107	AB002287	Wolbachia sp. DNA for GroES protein homolog, GroEL protein homolog, partial cds	6.60E-01
108	X16349	Human gene for sex hormone-binding globulin (SHBG)	2.40E-02
109	Z97349	Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 3-06, complete sequence	6.50E-01
110	L06898	Actinomyces viscosus sialidase (nanH) gene, complete cds.	2.20E-01
111	J03818	Rhesus monkey psi-eta-globin gene intergenic region, with Alu repeats.	2.10E-01
112	U66708	Vibrio parahaemolyticus ClpX-like protein (clpX) gene, partial cds, and lon protease (lonS) gene, complete cds	1.90E+00
113	AF078164.2	Homo sapiens Ku70-binding protein (KUB3) mRNA, partial cds	e-174
114	AJ010642	Drosophila melanogaster mRNA for Dof protein, transcript I, partial	1.90E+00
115	AF039096	Diadasia martialis cytochrome oxidase I (CO1) gene, mitochondrial gene encoding mitochondrial protein, partial cds	8.10E-01
116	L44593	Bacteriophage BK5-T ORF'410, 3' end pf cds, 20 ORFs, repressor protein, and Cro repressor protein genes, complete cds, ORF70' gene, 5' end of cds.	2.30E-01
117	U71249	Drosophila virilis cecropin 1 (Cec1), cecropin 2 complete cds and cecropin, pseudogene, exon 1	0.22
118	AL049223.1	Homo sapiens mRNA; cDNA DKFZp564L1916 (from clone DKFZp564L1916)	e-161
119	D13158.2	Bacillus sp. gene for thermostable alkaline protease, complete cds	0.69
120	M36287	S.cerevisiae alpha-aminoadipate reductase (LYS2) gene, complete cds.	6.70E-01
121	AF083457.1	Equus caballus microsatellite COR014 sequence	8.00E-03
122	X66015	T.aestivum mRNA 3 for cathepsin B (2557)	8.00E-03
123	U42767	Drosophila melanogaster leucine-rich repeat/Ig transmembrane protein KEK1 precursor (kek1) mRNA, complete cds	1.90E+00
124	X06670	Yeast NUC1 gene for mitochondrial nuclease	7.10E-02
125	Z49613	S.cerevisiae chromosome X reading frame ORF YJR113c	6.50E-01
126	U00038	Caenorhabditis elegans cosmid T21D11	2.20E-01
127	M60177	Escherichia coli enterobactin (entF) gene, complete cds.	6.50E-01
128	Z84506	H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA28B10	2.10E-01
129	J00334	Monkey (rhesus) delta-globin pseudogene; 5' flank and exons 1 & 2.	2.10E-01

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
130	NM_002844.1	Homo sapiens protein tyrosine phosphatase, receptor type, K (PTPRK) mRNA phosphatase mRNA, complete cds	7.00E-03
131	U86751	Human nucleolar fibrillar center protein (ASE-1) mRNA, complete cds	8.00E-03
132	D63735	Distolasterias nipon DNA for 16S rRNA, partial sequence	3.00E-03
133	D13469	M.hyopneumoniae genome, repeated DNA sequence	7.60E-02
134	Z15030	H.sapiens gene for ventricular myosin light chain 2 >gi|340286|gb|L01652|HUMVMLC Human ventricular myosin light chain 2 gene, seven exons.	7.60E-02
135	AL035426.2	Human DNA sequence from clone 370N13 on chromosome Xq25-26.3. Contains an exon of the GRIA3 gene for glutamate receptor, ionotrophic, AMPA 3. Contains ESTs, complete sequence [Homo sapiens]	2.20E-01
136	U61420	Human myosin VIIa (MYO7A) gene, exons 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14	3.00E-03
137	AF155117.1	Homo sapiens NY-REN-62 antigen mRNA, partial cds	e-142
138	AE001129	Borrelia burgdorferi (section 15 of 70) of the complete genome	8.80E-02
139	AB014528	Homo sapiens mRNA for KIAA0628 protein, complete cds	2.00E-39
140	X85980	H.sapiens serine hydroxymethyltransferase pseudogene	2.40E-02
141	D16474	Human mRNA, Xq terminal portion	9.00E-04
142	NM_005180.1	Homo sapiens murine leukemia viral (bmi-1) oncogene homolog (BMI1) mRNA	8.00E-04
143	U78193	Borrelia burgdorferi tuf-s10 operon: elongation factor (tuf), ribosomal proteins S10 (rpsJ), L3 (rplC), L4 (rplD), L23 (rplW), L2 (rplB), S19 (rpsS), and L22 cds	1.00E-03
144	AF052168	Homo sapiens clone 24762 mRNA sequence	6.60E-01
145	NM_001863.1	Homo sapiens cytochrome c oxidase subunit VIb mRNA, complete sequence	9.00E-05
146	AB010273.1	Homo sapiens pshsp47 gene, complete cds	1.9
147	AL109729.1	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 123453	1E-81
148	U71187.1	Human cholesteryl ester transfer protein (CETP) gene, partial cds and promoter region	0.023
149	X90761	Homo sapiens hHa2 gene	0.0003
150	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.028
151	AL049300.1	Homo sapiens mRNA; cDNA DKFZp564P063 (from clone DKFZp564P063)	0.00001

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
152	AP000273.1	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region, clone:f80G10, complete sequence	0.003
153	NM_001277.1	Homo sapiens choline kinase (CHK) mRNA kinase	0.00003
154	AC001036	Homo sapiens (subclone 2_f7 from P1 H48) DNA sequence	0.002
155	U45432	Human ETV6 gene, promoter region and partial cds	0.008
156	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.25
157	X60175	D.silvestris clone U28T2 non-LTR retrotransposon DNA (3778 bp)	0.66
158	X93334	H.sapiens mitochondrial DNA, complete genome	0.00001
159	J04838	Human apolipoprotein B (APOB) gene, exons 21, 22 and 23.	0.000001
160	M94631	Hylobates lar (clone LambdaGialphaG1) 3'alpha1Alu1 D, 3'alpha1Alu1 E and 3'alpha1Alu1 F Alu repeat regions.	0.000003
161	Z81524	Caenorhabditis elegans cosmid F32H5, complete sequence [Caenorhabditis elegans]	0.71
162	U42364	Sus scrofa centromere-specific repeat, T32M clone, Mc2 satellite DNA amplified from S0048 primer set.	0.23
163	AJ238233.1	Homo sapiens RPC62 gene for RNA polymerase III subunit, exon 13	1E-35
165	AF030697	Homo sapiens semaphorin L (SEMAL) gene, partial cds	0.00000002
166	Y08639	H.sapiens mRNA for nuclear orphan receptor ROR-beta	0.092
167	X71934	H.sapiens XB gene for tenascin-X, repeat XIII	0.0001
168	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.03
169	M65243	Synthetic mRNA leader sequence UTK.	0.083
170	NM_000999.1	Homo sapiens ribosomal protein L38 (RPL38) mRNA >gi|407422|emb|Z26876|HSRPL38 H.sapiens gene for ribosomal protein L38	2E-09
171	X73501	H.sapiens gene for cytokeratin 20	3E-13
172	M17374	X.laevis beta-globin mRNA, 5' UTR.	9.00E-03
173	Z24233	H. sapiens (D12S352) DNA segment containing	2E-11
174	NM_003033.1	Homo sapiens sialyltransferase 4A mRNA >gi|410225|gb|L13972|HUMSIAT Homo sapiens beta-galactoside alpha-2,3-sialyltransferase (SIAT4A) mRNA, complete cds	7.00E-13
175	AF039652	Homo sapiens ribonuclease H type II mRNA, complete cds	9E-88
176	Z23435	H. sapiens (D1S414) DNA segment containing (CA) repeat; clone AFM179xg5; single read	0.001
177	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.009
178	AF128535.1	Mus musculus cytoplasmic phosphoprotein PACSIN2 mRNA, complete cds	2E-20
179	U19358	Saccharomyces cerevisiae dnaJ homolog Hlj1p	3.00E-14

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
180	AB011139	Homo sapiens mRNA for KIAA0567 protein, partial cds	4.00E-16
181	AF154851.1	Salvelinus alpinus mitochondrion complete genome	2.20E-01
182	AB028980.1	Homo sapiens mRNA for KIAA1057 protein, partial cds	4.00E-38
183	U39178	Human phosphodiesterase (PDEA) gene, intron 16, 3' end	1E-16
184	U02455	Cloning vector rpDR2, complete sequence.	6.00E-19
185	NM_006048.1	Homo sapiens clone 686 protein (KIAA0684) mRNA >gi|4104975|gb|AF043117|AF043117 Homo sapiens ubiquitin-fusion degradation protein 2 (UFD2) mRNA, complete cds	2.00E-64
186	AF002644	Limulus polyphemus cytochrome oxidase II complete sequence, ATP synthase 8 (ATPase 8) gene, complete cds, and ATP synthase 6 (ATPase 6) gene, partial cds, mitochond...	2.40E-02
187	Z58806	H.sapiens CpG island DNA genomic Mse1 fragment, clone 50f4, forward read cpg50f4.ft1a	6.00E-20
188	U58736	Caenorhabditis elegans cosmid EGAP7.	8.00E-03
189	V01270	R.norvegicus genes for 18S, 5.8S, and 28S ribosomal RNAs	6.00E-21
190	L42098	Homo sapiens (subclone 5_c7 from P1 H22) DNA sequence.	9.00E-14
191	Z63236	H.sapiens CpG island DNA genomic Mse1 fragment, clone 7f5, forward read cpg7f5.ft1d	2.00E-21
192	AF145957.2	Mus musculus groucho-related gene product	1.00E-57
193	NM_003193.1	Homo sapiens tubulin-specific chaperone e tubulin-folding cofactor E mRNA, complete cds	2.00E-23
194	U65980	Borrelia hermsii 38 kDa lipoprotein Gpd gene, complete cds	2.00E+00
195	U49974	Human mariner2 transposable element, complete consensus sequence	4.00E-28
196	Z47053	Human microsatellite DNA sequence	5E-29
197	X80424	M.musculus tex23 mRNA (5'region)	1.00E-27
198	U75467	Drosophila melanogaster Rga and Atu genes, complete cds	4.00E-28
199	U43077	Human CDC37 homolog mRNA, complete cds	1.00E-28
200	NM_000436.1	Homo sapiens 3-oxoacid CoA transferase mRNA >gi|1519051|gb|U62961|HSU62961 Human succinyl CoA:3-oxoacid CoA transferase precursor (OXCT) mRNA, complete cds	2.00E-29
201	NM_003979.1	Homo sapiens retinoic acid induced 3 (RAI3) mRNA >gi|4063889|gb|AF095448|AF095448 Homo sapiens putative G protein-coupled receptor (RAIG1) mRNA, complete cds	e-158
202	Z22466	H.sapiens DNA sequence	5E-30

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
203	X15880	Human mRNA for collagen VI alpha-1 C-terminal globular domain	7.00E-33
204	U47322	Cloning vector DNA, complete sequence.	8E-34
205	Z55306	H.sapiens CpG island DNA genomic Mse1 fragment, clone 32a6, forward read cpg32a6.ft1a	2E-20
206	AC005190	Homo sapiens PAC clone DJ1152D16 from Xq23, complete sequence [Homo sapiens]	1.00E-26
207	Z56833	H.sapiens CpG island DNA genomic Mse1 fragment, clone 14e3, reverse read cpg14e3.rt1b	7.00E-11
208	D38101	Rat rCACN4A mRNA for L-type voltage-dependent calcium channel alpha 1 subunit, complete cds	2.40E-02
209	Z56833	H.sapiens CpG island DNA genomic Mse1 fragment, clone 14e3, reverse read cpg14e3.rt1b	6.00E-11
212	NM_002004.1	Homo sapiens farnesyl diphosphate synthase dimethylallyltranstransferase, geranyltranstransferase) for KIAA0003 gene, complete cds	2.00E-43
213	AB023234.1	Homo sapiens mRNA for KIAA1017 protein, complete cds	e-172
214	NM_003492.1	Homo sapiens ITBA1 gene (ITBA1) mRNA protein	1.00E-49
215	X52994	Sheep mRNA for CD3 gamma subunit (partial)	5.00E-08
216	AF059650	Homo sapiens histone deacetylase 3 (HDAC3) gene, complete cds	6.80E-01
217	U49046	Mus musculus zinc finger protein (Zfp64) mRNA, complete cds	3.00E-55
218	NM_003488.1	Homo sapiens A kinase anchor protein, 149kD mRNA for kinase A anchor protein	3.00E-21
219	J03764	Human, plasminogen activator inhibitor-1 gene, exons 2 to 9.	3.00E-26
220	Y16675	Homo sapiens mRNA for aflatoxin B1-aldehyde reductase	8.00E-03
221	AF085715	Mus musculus homeobox protein SPX1 mRNA, complete cds	2.10E-01
222	X76968	Loligo forbesi mRNA for phosphatidylinositol-specific phospholipase C	1.9
223	X55741	H.sapiens FKBP cDNA	2.00E-65
224	X85060	B.taurus cosmid-derived microsatellite DNA	3.00E-76
225	AJ001119	Bos taurus mRNA for Rab5 GDP/GTP exchange factor, Rabex5	3E-79
226	D63850	Mus musculus mRNA for hepatoma-derived growth factor, complete cds, strain:BALB/c	e-102
227	AB018344.1	Homo sapiens mRNA for KIAA0801 protein, complete cds	e-169
228	X81058	M.musculus tex261 mRNA	e-112

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
229	AF041853	Homo sapiens kinesin family member protein KIF3A mRNA, complete cds	e-162
234	AB020395	Taenia taeniaeformis mitochondrial DNA for large subunit ribosomal RNA, partial sequence	1.90E+00
235	AF120325.1	Cricetulus griseus class I beta tubulin gene, complete cds	1.80E+00
237	AF084259	Mus musculus bromodomain-containing protein BP75 mRNA, complete cds	0.64
239	M77820	Xenopus laevis fibronectin mRNA, complete cds.	2
241	AB007930	Homo sapiens mRNA for KIAA0461 perotein, partial cds	e-178
242	D87463	Human mRNA for KIAA0273 gene, complete cds	2
244	AF094519	Mus musculus diaphanous-related formin (Dia2) mRNA, complete cds	e-143
246	Z69708	Human DNA sequence from cosmid L241B9, Huntington's Disease Region, chromosome 4p16.3 contains polymorphic VNTR pYNZ32	2
247	AF156102.1	Homo sapiens ELL complex EAP30 subunit mRNA, complete cds	e-169
248	L36592	Homo sapiens kidney epithelial sodium channel gamma subunit (gamma hENaC) mRNA, complete cds.	0.63
250	AF039945	Homo sapiens synaptojanin 2B mRNA, partial cds	2.1
252	U29487	Caenorhabditis elegans cosmid C09C7	0.71
253	NM_003794.1	Homo sapiens sorting nexin 4 (SNX4) mRNA nexin 4 mRNA, complete cds	e-151
255	AE001267	Treponema pallidum section 83 of 87 of the complete genome	6.70E-01
256	NM_005686.1	Homo sapiens SRY (sex determining region Y)-box 13 (SOX13) mRNA >gi|4323170|gb|AF098915|AF098915 Homo sapiens type 1 diabetes autoantigen ICA12 mRNA, complete cds	0.23
257	Z49373	S.cerevisiae chromosome X reading frame ORF YJL098w	2
258	AF125392.1	Homo sapiens insulin induced protein 2 mRNA, complete cds	8.00E-96
260	X07618	Human mRNA for cytochrome P450 db1 variant a	6.90E-01
261	AB022161.1	Mus musculus Cctq gene for chaperonin containing TCP-1 theta subunit, complete cds	0.7
262	AF001794	Mus musculus Treacher Collins Syndrome protein	0.69
263	AF119362.1	Mus musculus strain 129/SvJ mast cell protease 8 (Mcpt8) gene, complete cds	0.22
264	AE001395	Plasmodium falciparum chromosome 2, section 32 of 73 of the complete sequence	6.80E-01
265	U19775	Human MAP kinase Mxi2 (MXI2) mRNA, complete cds	2.10E+00

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
266	AJ131389	Homo sapiens mRNA for PEX3 protein, partial	e-171
267	M25779	S.cerevisiae SEC59 gene, complete cds.	1.90E+00
268	AF009953	Glycine max 35 kDa seed maturation protein	0.66
269	Z35284	H.sapiens mRNA for MDR3 P-glycoprotein	2.40E-02
270	AF050052	Pleurocera prasinatum strain 12B-1 16S ribosomal RNA gene, mitochondrial gene for mitochondrial RNA, partial sequence	6.60E-01
271	AF043494	Pinus strobus microsatellite RPS3 repeat region	6.60E-01
272	NM_001388.1	Homo sapiens developmentally regulated GTP-binding protein 2 (DRG2) mRNA GTP-binding protein	0.65
273	NM_000242.1	Homo sapiens mannose-binding lectin, soluble mannose-binding protein C	6.90E-01
274	S82740	NPM/ALK=fusion gene {translocation breakpoint}	7.10E-01
275	AB006621	Homo sapiens mRNA for KIAA0283 gene, partial cds	1.90E+00
276	AB023162.1	Homo sapiens mRNA for KIAA0945 protein, complete cds	e-169
277	AF124490.1	Homo sapiens ARF GTPase-activating protein GIT1 mRNA, complete cds	e-173
278	L29454	Mouse fibrillin (Fbn-1) mRNA, complete cds.	0.64
279	AB018264.1	Homo sapiens mRNA for KIAA0721 protein, partial cds	e-148
280	AF046000	Mus musculus rod cGMP phosphodiesterase delta subunit (Pde6d) gene, complete cds	0.52
281	X86791	S.scrofa beta-globin gene	0.37
282	AL050368.1	Homo sapiens mRNA; cDNA DKFZp566A1124	2.1
283	AF174426.1	Acholeplasma laidlawii DNA topoisomerase IV ParE subunit (parE) and DNA topoisomerase IV ParC subunit (parC) genes, partial cds	2.1
284	AJ009770	Homo sapiens mRNA for putative transcription factor, partial	e-165
285	U89992	Mus musculus lymphocyte-specific adaptor protein Lnk (Lnk) mRNA, complete cds	0.23
287	AF132479	Mus musculus Ese2L protein mRNA, complete cds	0.7
288	X76753.2	Homo sapiens HG 5-HTT gene for serotonin transporter, exon 1	2.1
289	D64033	Oryzias latipes DNA for transferrin, complete cds	0.23
290	D29985	Bacillus subtilis wapA and orf genes for wall-associated protein and hypothetical proteins	0.68
291	AL049442.1	Homo sapiens mRNA; cDNA DKFZp586N1720 (from clone DKFZp586N1720)	e-166

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
292	S64907	cgs2+=cyclic AMP dependent protein kinase regulatory subunit homolog [Schizosaccharomyces pombe=fission yeast, Genomic, 3596 nt]	0.68
293	AF027202	Bos taurus rod outer segment guanylate cyclase precursor (ROS-GC1) gene, exons 9 through 11	2
294	AB011540	Homo sapiens mRNA for MEGF7, partial cds	0.076
295	AL096842.1	Homo sapiens mRNA; cDNA DKFZp586D1519 (from clone DKFZp586D1519)	e-177
296	D78503.1	Mus musculus seizure-related mRNA, partial sequence	0.68
297	AF079557	Mus musculus poly(ADP-ribose) glycohydrolase	2
298	Z66316	H.sapiens CpG island DNA genomic Mse1 fragment, clone 8a6, forward read cpg8a6.ft1f	0.22
299	U86453	Human phosphatidylinositol 3-kinase catalytic subunit p110delta mRNA, complete cds	2.1
300	AJ000467.1	Crocidura russula partial mitochondrial cytb gene >gi|3319900|emb|AJ000468.1|CRAJ468 Crocidura russula partial mitochondrial cytb gene	0.22
301	X63721	S.cerevisiae HEM12 gene for uroporphyrinogen decarboxylase	0.67
302	AJ005390.1	Homo sapiens SCNN1B gene, exons 9 and 10	0.23
303	X06150	Rat mRNA for glycine methyltransferase (EC 2.1.1.20)	0.22
304	X63771.1	Soybean Mosaic Virus gene for coat protein	2
305	AF113615.1	Homo sapiens FH1/FH2 domain-containing protein FHOS (FHOS) mRNA, complete cds	e-176
306	AF052193	Gallus gallus translation repressor mRNA, partial cds	0.66
307	D13903	Mouse mRNA for MPTPdelta (type A)	0.22
308	M77144	Human type II 3-beta hydroxysteroid dehydrogenase/ 5-delta - 4-delta isomerase gene, complete cds.	0.22
309	AJ236656	Homo sapiens chromosome 22 CpG island DNA, genomic Mse1 fragment, clone 22CGIB49B8 , complete read	0.66
310	M84732	Plasmodium yoelii sporozoite surface protein 2 gene	0.22
311	X83433	O.sativa mRNA for lipid transfer protein, b21	0.66
312	AP000145.1	Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region, clone B2291C14-R44F3, segment 10/10, complete sequence	0.0000004
313	AB025570.1	Equus caballus CgA mRNA for chromogranin A, complete cds	0.22
314	AF006482	Mus musculus nucleoside triphosphatase	0.69
315	AF141308.1	Homo sapiens polyamine modulated factor-1	0.65

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
316	AF092945	Charybdis feriatus molt-inhibiting hormone	0.22
317	D90773	E.coli genomic DNA, Kohara clone #262(30.3-30.5 min.)	1.9
318	D26077	Mouse mRNA for KIF3B protein, complete cds	0.21
319	U68036	Streptomyces coelicolor bldKA, bldKB, bldKC, and bldkD genes, complete cds, and bldkE gene, partial cds	0.64
321	X75563	S.oleracea mRNA (omp24) for chloroplast outer envelope 24 kD protein	0.68
322	AB006628	Homo sapiens mRNA for KIAA0290 gene, partial cds	0.21
323	AJ238878.1	Haloferax volcanii ORF1, strain WR340	0.21
324	U39696	Mycoplasma genitalium section 18 of 51 of the complete genome	0.21
326	AL031590	Human DNA sequence from clone 232D4 on chromosome 22q13.1 Contains GSS, complete sequence [Homo sapiens]	0.67
327	AF103731.1	Homo sapiens putative glycolipid transfer protein mRNA, complete cds	e-168
328	U88984	Mus musculus NIK mRNA, complete cds	0.22
329	AJ006031	Mus musculus IHABP gene, promoter	2E-40
330	AL049953.1	Homo sapiens mRNA; cDNA DKFZp564P0622	6E-52
331	L49144	Homo sapiens neuroendocrine-specific protein	0.81
332	U25810	Bos taurus lysozyme (LZ) gene, complete cds	0.000004
333	AF092681	Exema neglecta haplotype 188 cytochrome oxidase I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds	0.77
334	AE000966	Archaeoglobus fulgidus section 141 of 172 of the complete genome	2.1
335	AF091234	Mus musculus putative transcription factor mRNA, complete cds	4E-90
336	M25702	Human thyroid peroxidase (TPO) gene, exon 2.	0.078
337	Z70029	B.vulgaris mitochondrial DNA, RAPD fragment	0.075
338	AF072432	Dictyostelium discoideum gp63 homolog mRNA, complete cds	0.69
339	M13241	Human N-myc gene, exons 2 and 3.	0.074
340	X07703	Chironomus tentans Balbiani ring gene BR6 3'-end	0.076
341	X57564	A.rusticana mRNA for neutral peroxidase	0.077
342	Z74084	S.cerevisiae chromosome IV reading frame ORF YDL036c	2
343	U30248	Caenorhabditis elegans transcription factor E12/47 homolog gene, complete cds	2.1
344	Y09396	C.annuum mRNA for CDC48p-like protein	2
345	Z92835	Caenorhabditis elegans cosmid H19N07, complete sequence [Caenorhabditis elegans]	0.68

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
346	M92844	Homo sapiens zinc finger transcriptional regulator (GOS24) gene, complete cds	2
347	X54111	Treponema pallidum GroEL gene and gene encoding putative enol-pyruvyltransferase	0.22
348	Z71419	S.cerevisiae chromosome XIV reading frame ORF YNL143c	0.64
349	AL034486	S.pombe chromosome I cosmid c2H10	1.9
350	NM_000127.1	Homo sapiens exostoses (multiple) 1 (EXT1) mRNA	2
351	U61997	Zea mays B chromosome centromere repeat K11 sequence	0.074
352	AB018255.1	Homo sapiens mRNA for KIAA0712 protein, complete cds	7E-43
353	S60289	LeB4=legumin {5' region} [Vicia faba, Genomic, 1222 nt]	0.072
354	AE001391	Plasmodium falciparum chromosome 2, section 28 of 73 of the complete sequence	0.24
357	U67478	Methanococcus jannaschii section 20 of 150 of the complete genome	0.068
358	AE001146	Borrelia burgdorferi (section 32 of 70) of the complete genome	1.9
359	U46542	Streptococcus crista HmpA gene, partial cds, putative adhesin/ABC transport system protein (scbA) gene, complete cds	0.073
360	Z36067	S.cerevisiae chromosome II reading frame ORF YBR198c	1.3
361	U78684	Teucridium parvifolium NADH dehydrogenase (ndhF) gene, chloroplast gene encoding chloroplast protein, partial cds	0.29
362	AF037332	Homo sapiens Eph-like receptor tyrosine kinase hEphB1b (EphB1) mRNA, complete cds	0.26
363	U54469	Drosophila melanogaster eukaryotic initiation factors 4E-I and 4E-II (eIF4E) gene, complete cds.	0.24
364	U20611	Mus musculus thioredoxin-dependent peroxide reductase (tpx) mRNA, complete cds.	0.027
365	L34542	Rattus norvegicus non-receptor protein kinase	0.7
366	Y14993	Schizosaccharomyces pombe gut2 gene	0.23
367	AL008983	Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 3-54, complete sequence	0.025
368	AL080129.1	Homo sapiens mRNA; cDNA DKFZp434D193 (from clone DKFZp434D193)	e-100
369	AF100304	Caenorhabditis elegans cosmid W07B3	0.65
370	AF039527	Bacillus stearothermophilus limonene hydroxylase (pOT435) gene, complete cds	0.22
371	AP000258.1	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region, clone:Q89A6, complete sequence	0.00001

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
372	AF082519	Entamoeba histolytica 70 kDa heat shock protein Hsp70-Bip precursor (BiP) gene, complete cds	0.0009
373	M38224	T.brucei procyclic acidic repetitive protein	1.9
374	Z70720	S.pombe chromosome I cosmid c1B9	0.65
375	AF069532	Homo sapiens CDP-diacylglycerol synthase 2	5E-20
376	X97570	Z.mays dek34 gene	0.22
377	NM_004652.1	Homo sapiens ubiquitin specific protease 9, X chromosome (Drosophila fat facets related) (USP9X) mRNA ubiquitin hydrolase	0.023
378	AJ223578	Branchiostoma lanceolatum mRNA for intermediate filament protein C2	0.024
379	D63523.1	Dictyostelium discoideum mitochondrial genes for ribosomal proteins, complete and partial cds	0.22
380	L35528	Mus musculus manganese superoxide dismutase	0.074
381	NM_004267.1	Homo sapiens carbohydrate (chondroitin 6/keratan) sulfotransferase 2 (CHST2) mRNA mRNA for N-acetylglucosamine-6-O-sulfotransferas e mRNA for long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST), complete cds	0.003
382	Z81507	Caenorhabditis elegans cosmid F18A11, complete sequence [Caenorhabditis elegans]	1.9
383	AF072506.2	Homo sapiens endogenous retrovirus W envelope protein precursor mRNA, complete cds	0.75
384	U39730	Mycoplasma genitalium cdsA, frr, hsdS, smbA, tsf genes from bases 539564 to 546816 (section 52 of 56) of the complete genome	0.009
385	L20296	Saccharomyces cerevisiae (chromosome II) ARO4-homologue (YBR1701), YBR1702, YBR1703, 30S ribosomal protein-homologue (YBR1704) and pseudoprotease-homologue	2
386	D50500	Mouse mRNA for Rab 11, partial sequence	0.22
387	X52263	C.tentans balbiani ring 3 (BR3) gene	2
388	M68998	Human alpha-1 type XIII collagen (COL13A1) gene, exon 1.	0.008
389	L40608	Plasmodium falciparum (strain Dd2) variant-specific surface protein (var-1) gene, complete cds.	2
390	D12688	Mouse P-cadherin gene, exon 1 and 2	2
391	AE001150	Borrelia burgdorferi (section 36 of 70) of the complete genome	0.008
392	X80852	M.musculus gene for liver type phosphofructokinase	0.073
393	AF115849.1	Trichomonas vaginalis pre-mRNA processing 8 protein homolog PRP8 (PRP8) gene, complete cds	2

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
394	L31848	Homo sapiens serine/threonine kinase receptor 2	0.069
395	AJ003222	Borrelia burgdorferi flgK, flbF, thdF, gidA, gidB, moxR, orf1, orf2, orf3, orf4 and orf5 genes	0.006
396	AB028958.1	Homo sapiens mRNA for KIAA1035 protein, partial cds	0.055
397	AF155110.1	Homo sapiens NY-REN-45 antigen mRNA, complete cds	0.07
398	M24842	Human keratin 18 (K18) gene, complete cds.	e-142
399	AL050074.1	Homo sapiens mRNA; cDNA DKFZp566F1946	e-171
401	D13469	M.hyopneumoniae genome, repeated DNA sequence	0.003
402	U67510	Methanococcus jannaschii section 52 of 150 of the complete genome	0.074
403	AB029343.1	Homo sapiens HCR (a-helix coiled-coil rod homologue) gene, complete cds	0.21
404	L43391	Homo sapiens (subclone 5_g12 from P1 H16) DNA sequence.	0.7
405	AF016864.1	Orpinomyces sp. PC-2 beta-glucosidase (bgl1) mRNA, complete cds	0.22
406	L31848	Homo sapiens serine/threonine kinase receptor 2	0.072
407	X65521	K.lactis centromere 2 (KlCEN2) DNA	0.024
408	U56221	HIV-1 clone 13Pb9-4 from Seattle, envelope glycoprotein, V3-V5 region (env) gene, partial cds	0.22
409	AF157816.1	Homo sapiens cAMP specific phosphodiesterase products, complete cds	2E-11
410	AF131748	Homo sapiens clone 25191 GTP-specific succinyl-CoA synthetase beta subunit (SCS) mRNA sequence, partial cds	0.23
411	AF034783	Synthetic helper virus genomic sequence fragment	2
412	AF035606	Homo sapiens calcium binding protein (ALG-2) mRNA, complete cds	0.000004
413	L07944	Plasmodium falciparum secreted polymorphic antigen gene, complete cds	0.001
414	AE001418	Plasmodium falciparum chromosome 2, section 55 of 73 of the complete sequence	0.026
415	Z68886	Human DNA sequence from cosmid L21F12, Huntington's Disease Region, chromosome 4p16.3	7E-12
416	X82192	H.sapiens EST mRNA (G5)	0.23
417	NM_004998.1	Homo sapiens myosin IC (MYO1C) mRNA complete cds.	0.0001
418	U55042	Bos taurus myosin X, complete cds	0.2

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
419	AF026069.1	Homo sapiens phosphomevalonate kinase	0.66
420	AL080128.1	Homo sapiens mRNA; cDNA DKFZp434C153 (from clone DKFZp434C153)	0.62
421	S75476	PGK1=phosphoglycerate kinase 1 {3' nuclease-sensitive region} [human, Genomic, 3571 nt]	0.00003
422	M57682	Rat brain calcium channel alpha-1 subunit mRNA, complete cds.	0.0001
423	AB023053.1	Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, clone:53L9, complete sequence	0.074
424	U74651	Human DNA polymerase gamma (polg) gene, promoter region and partial cds	7E-11
426	X86336	H.sapiens C7 gene, exon 9	0.026
427	AB000931.2	Homo sapiens FUT2 gene, intron 1, complete sequence	0.0003
428	U20365	Mus musculus smooth muscle gamma-actin gene, complete cds	0.0003
429	AF136745.1	Homo sapiens diacylglycerol kinase epsilon gene, promoter and partial cds	0.0000001
430	X04249	Human gene for small cytoplasmic 7SL RNA (7L30.1) pseudogene	0.000001
431	AB029016.1	Homo sapiens mRNA for KIAA1093 protein, partial cds	0.00000005
432	AE001421	Plasmodium falciparum chromosome 2, section 58 of 73 of the complete sequence	0.001
433	AB023189.1	Homo sapiens mRNA for KIAA0972 protein, complete cds	0.003
434	U68061	Human MUC2 gene, promoter region	0.000001
435	NM_005971.1	Homo sapiens phospholemman-like, expressed in breast tumors, 8kD (PLML) mRNA protein	5E-09
436	AC001050	Homo sapiens (subclone 3_e9 from P1 H55) DNA sequence	5E-09
437	AF151843.1	Homo sapiens CGI-85 protein mRNA, complete cds	1E-35
438	U26447	Human natural resistance-associated macrophage protein (NRAMP1) gene, 3' region	6E-10
439	Z95309	Caenorhabditis elegans cosmid H36L18, complete sequence [Caenorhabditis elegans]	2
440	AF144622.1	Homo sapiens beta-catenin gene, intron 2 and partial cds	2
441	J04990	Human cathepsin G gene, complete cds.	0.0000001
442	U22657	Mus musculus genomic locus related to cellular morphology.	0.076

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
443	AP000262.1	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region, clone:S680, complete sequence	2E-12
444	AF115549.2	Homo sapiens Wiskott-Aldrich Syndrome protein flanking region	6E-21
445	M55409	Homo sapiens pancreatic tumor-related protein mRNA, partial cds	8E-13
446	NM_006530.1	Homo sapiens Glioma-amplified sequence-41 GAS41 protein mRNA, complete cds	e-154
447	U30261	Schistosoma mansoni G protein beta subunit-like protein trans-spliced mRNA, complete cds	3E-14
448	AF132966.1	Homo sapiens CGI-32 protein mRNA, complete cds	e-169
449	M15205	Human thymidine kinase gene, complete cds, with clustered Alu repeats in the introns.	1E-14
450	X92565	C.elegans mRNA for LIN-2B protein	0.0000001
451	NM_006466.1	Homo sapiens polymerase (RNA) III (DNA directed) (39kD) (RPC39) mRNA subunit (RPC39) mRNA, complete cds	3E-15
452	AF086460	Homo sapiens full length insert cDNA clone ZD85A02	e-117
453	L35664	Homo sapiens (subclone H8 8_f5 from P1 35 H5 C8) DNA sequence.	2E-10
454	X69951	H.sapiens gene for casein kinase II alpha subunit	2E-20
455	AB007930	Homo sapiens mRNA for KIAA0461 perotein, partial cds	e-177
456	L81840	Homo sapiens (subclone 1_f8 from P1 H43) DNA sequence	1E-27
457	X94354	H.sapiens DNA for Cone cGMP-PDE gene	4E-17
458	AB024291.1	Zea mays ZmRR2 mRNA, complete cds	0.025
459	Y16790	Homo sapiens hHa4 gene, complete CDS	0.66
460	U67209	Human clone HS2.10 Alu-Ya5 sequence	2E-19
461	M30951	Gorilla 28S ribosomal RNA gene fragment.	5E-20
462	AF029062	Homo sapiens DEAD-box protein (BAT1) gene, partial cds	1E-18
463	AB014601	Homo sapiens mRNA for KIAA0701 protein, partial cds	1E-14
464	M30950	Chimpanzee 28S ribosomal RNA gene fragment.	6E-21
465	AB005619	Gallus gallus mRNA for chromobox protein	3E-26
466	AF070657	Homo sapiens glutathione S-transferase subunit 13 homolog mRNA, complete cds	2E-54
467	M58775	Polaribacter glomeratus 16S ribosomal RNA	2.1
468	AJ000992.1	Dictyostelium discoideum gdt1 gene	0.67
469	AB014589	Homo sapiens mRNA for KIAA0689 protein, partial cds	e-158
470	Z63830	H.sapiens CpG island DNA genomic Mse1 fragment, clone 90h2, reverse read cpg90h2.rt1a	3E-26

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
471	NM_002273.1	Homo sapiens keratin 8 (KRT8) mRNA keratin 8	e-120
472	AF116910.1	Homo sapiens clone HAW100 putative ribonuclease III mRNA, complete cds	e-173
473	AF131739	Homo sapiens clone 25189 mRNA sequence, complete cds	e-124
474	AF100615.1	Homo sapiens chromosome 15 MRG15 protein	7E-74
475	AB019490.1	Homo sapiens IDN4-GGTR7 mRNA, partial cds	e-156
476	L20941	Human ferritin heavy chain mRNA, complete cds.	1E-27
477	AF088022	Homo sapiens full length insert cDNA clone ZC18H06	5E-30
478	L06845	Human cysteinyl-tRNA synthetase mRNA, partial cds.	1E-39
479	AB014542	Homo sapiens mRNA for KIAA0642 protein, partial cds	2E-54
480	L77890	Homo sapiens excision repair protein ERCC4 mRNA, complete cds, clone cer4-40	2E-30
481	L32838	Mouse germline interleukin 1 receptor antagonist	0.076
482	NM_004537.1	Homo sapiens nucleosome assembly protein 1-like 1 (NAP1L1) mRNA >gi|189066|gb|M86667|HUMNAP H.sapiens NAP (nucleosome assembly protein) mRNA, complete cds	e-123
483	U85258	Human estrogen related receptor alpha (ESTRRA) pseudogene	8E-34
484	U79656	Human Treacher Collins syndrome (TCOF1) gene, exon 21	8E-34
485	AF067864.1	Homo sapiens transferrin receptor 2 alpha	4E-91
486	X03100	Human HLA-SB(DP) alpha gene	1E-16
487	AF013277	Bombyx mori topoisomerase II (TOPOII) mRNA, complete cds	0.23
488	U46068	Mus musculus von Ebner minor salivary gland protein mRNA, complete cds.	1E-35
489	U67563	Methanococcus jannaschii section 105 of 150 of the complete genome	1E-35
490	AB016492.1	Homo sapiens hJTB gene, complete cds	e-118
491	X98176	H.sapiens mRNA for MACH-beta-1 protein	1E-36
492	AF049613	Homo sapiens huntingtin interacting protein HYPK mRNA, partial cds	7E-22
493	AF039690.1	Homo sapiens antigen NY-CO-8 (NY-CO-8) mRNA, partial cds	1E-37
494	NM_001003.1	Homo sapiens ribosomal protein, large, P1 ribosomal phosphoprotein P1 mRNA, complete cds.	4E-38

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
495	U34305	Shigella sonnei form I operon ORF protein genes, complete cds, insertion sequence IS630 protein gene, complete cds.	0.074
496	U61538	Human calcium-binding protein chp mRNA, complete cds	4E-38
497	AJ243512.1	Homo sapiens mRNA for Barx2 protein (Barx2 gene)	1E-46
498	AF077043.1	Homo sapiens 60S ribosomal protein L36 mRNA, complete cds	4E-59
499	Y14223	Homo sapiens BPI gene, exon 9	0.00001
500	X07425	Human gene for U 6 RNA	1E-35
501	U43508	Mus musculus RORgamma orphan nuclear receptor mRNA, complete cds	0.23
502	Z92541	Human DNA sequence from PAC 179I15, BRCA2 gene region chromosome 13q12-13 contains lactase-phlorizin hydrolase (LCT)	0.078
503	X57435	H.sapiens mRNA for transcription factor AP-4	0.26
504	X70154	Z.mays mRNA for b-32 protein, putative regulatory factor of zein expression (clone b-32.152)	2.1
505	AF069737	Xenopus laevis notchless (nle) mRNA, complete cds	2E-94
506	D63850	Mus musculus mRNA for hepatoma-derived growth factor, complete cds, strain:BALB/c	5E-50
507	NM_006295.1	Homo sapiens valyl-tRNA synthetase 1 (VARS1) mRNA	2E-50
508	Y16355	Homo sapiens mRNA for protein encoded by cxorf5 (71-7A) gene, alternatively spliced form	e-157
509	U67317	Cuphea wrightii beta-ketoacyl-ACP synthase II	0.68
510	NM_006571.1	Homo sapiens novel RGD-containing protein mRNA, complete cds	1E-56
511	NM_003574.1	Homo sapiens VAMP (vesicle-associated membrane protein)-associated protein A (33kD) (VAPA) mRNA, and translated products VAMP-associated protein of 33 kDa (VAP-33) mRNA, complete cds	e-129
512	NM_003431.1	Homo sapiens zinc finger protein 124 (HZF-16) HZF-16=Kruppel-related zinc finger gene homolog HEP-G2, mRNA, 2080 nt]	2E-60
513	J03798	Human autoantigen small nuclear ribonucleoprotein Sm-D mRNA, complete cds.	2E-72
514	AL049670.1	Human gene from PAC 69E11, chromosome 1	e-174
515	AB014603	Homo sapiens mRNA for KIAA0703 protein, complete cds	e-167
516	NM_000977.1	Homo sapiens ribosomal protein L13 (RPL13) mRNA >gi|29382|emb|X64707|HSBBC1 H.sapiens BBC1 mRNA	2E-63

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
517	Z55204	H.sapiens CpG island DNA genomic Mse1 fragment, clone 26c2, reverse read cpg26c2.rt1a	1E-28
518	AC002181	Homo sapiens (subclone 2_a12 from BAC H111) DNA sequence	0.001
519	NM_006371.1	Homo sapiens cartilage-associated protein sapiens mRNA for cartilage-associated protein (CASP)	e-171
520	AF102507.1	Homo sapiens fizzy-related protein mRNA, partial cds	e-153
521	U91561	Rattus norvegicus pyridoxine 5'-phosphate oxidase mRNA, complete cds	e-100
522	X56974	M.musculus mRNA for external transcribed spacer	e-163
523	AF060539	Mus musculus channel interacting PDZ domain protein mRNA, complete cds	e-138
524	AF071592	Homo sapiens kinesin superfamily motor KIF4 mRNA, complete cds	0
525	X68199	R.norvegicus MYR1 mRNA for myosin I heavy chain	e-128
526	NM_006693.1	Homo sapiens no arches-like (zebrafish) zinc finger protein (NAR) mRNA >gi|4098571|gb|U79569|HSU79569 Human no arches (nar) mRNA, complete cds	e-160
527	Z22818	Canis familiaris mRNA for Rab12 protein	e-159
529	AF077330	Mus musculus NEDD8-conjugating enzyme (Uba3) mRNA, complete cds	0.62
532	AF118268.1	Coprinus cinereus laccase 2 precursor (lcc2) gene, complete cds	2
533	AF118268.1	Coprinus cinereus laccase 2 precursor (lcc2) gene, complete cds	1.9
534	U33265	Coccidioides immitis complement fixation/chitinase antigen mRNA, complete cds	1.8
538	AF079867.1	Acomys cahirinus clone pAcah3 satellite sequence	1.8
539	NM_001324.1	Homo sapiens cleavage stimulation factor, 3' pre-RNA, subunit 1, 50kD (CSTF1) mRNA pZ50-19) cleavage stimulation factor 50kDa subunit, complete cds.	0.69
540	AB012265	Mus musculus mRNA for wizL, complete cds	0.64
541	Y18504.1	Homo sapiens X5L gene	e-151
542	AF034265	Gracilaria chilensis 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 25S ribosomal RNA gene, partial sequence	0.62
543	U16163	Mus musculus prolyl 4-hydroxylase alpha(II)-subunit mRNA, complete cds	0.62

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
544	U53004	Human GT335 gene, exons 1, 2, 3, and 4	0.61
545	Y13870.1	Homo sapiens mRNA containing (CAG)6 repeat, clone CZ-CAG 12	0.22
546	AF127950.1	Homo sapiens DNA polymerase epsilon catalytic subunit protein (POLE1) gene, exons 17, 18 and 19	0.21
547	AF071538	Homo sapiens Ets transcription factor PDEF	e-166
548	D63876	Human mRNA for KIAA0154 gene, partial cds	0.61
549	AE001326	Chlamydia trachomatis section 53 of 87 of the complete genome	2.3
550	Z58704	H.sapiens CpG island DNA genomic Mse1 fragment, clone 49b2, reverse read cpg49b2.rt1b	2.3
551	X78576	R.oryzae fumR gene	0.22
552	AB014740.1	Oryza sativa gypsy-type retrotransposon RIRE8A DNA, internal region, complete sequence	0.64
553	X78562	O.limosus hypoglycemic hormone mRNA CHAA,2409bp	0.21
554	X99719	S.enterica hsdM, hsdS & hsdR genes	1.9
555	U58513	Mus musculus Rho-associated, coiled-coil forming protein kinase p160 ROCK-2 mRNA, complete cds	1.9
556	Z83002	B.pagrosomi partial 28S rRNA gene	0.66
557	AL080223.1	Homo sapiens mRNA; cDNA DKFZp566H2446	e-150
558	AL080066.1	Homo sapiens mRNA; cDNA DKFZp564J142 (from clone DKFZp564J142)	0.00003
559	AF020424	Nicotiana tabacum glutamate decarboxylase isozyme 2 (NtGAD2) mRNA, complete cds	1.8
560	U32768	Haemophilus influenzae Rd section 83 of 163 of the complete genome	0.21
561	M10316	Plasmid pJD1 from Neisseria gonorrheae DNA, complete genome.	2
562	AB004272.1	Bos taurus mRNA for placenta growth factor precursor, complete cds	1.9
563	X05427	Drosophila ultrabithorax (Ubx) gene promoter region	1.9
564	M18729	S.pneumoniae mismatch repair protein (hexA) gene, complete cds.	0.21
565	M87060	Rattus rattus cardiac AE3 gene, exons 1-23.	0.086
566	M36662	Chicken alpha-1 collagen type III gene, 3' end.	0.083
567	AF135450.1	Sus scrofa SMCY (SMCY) gene, partial cds	0.081
568	Z34293	A.thaliana (CDNA4) myosin heavy chain mRNA	2.2
569	U83880	Rattus norvegicus glycerol-3-phosphate dehydrate dehydrogenase (mtGPDH) mRNA, 3'UTR	1E-59

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
570	AF133913.1	Mus musculus ARL-6 interacting protein-6	4E-79
571	AF077543.1	Caenorhabditis elegans cosmid H07I21	1.9
572	X77829	A.niger (N400) gsdA gene	0.07
573	X74765	H.sapiens CSK gene for protein tyrosine kinase	0.069
574	X63510	M.musculus CAML1 gene (exons 5-9)	0.62
575	U27319	Rattus norvegicus type I hexokinase (HKI) gene, promoter region and partial cds	0.61
576	L23863	Rat Skn1i mRNA.	0.068
577	Z26284.1	H.sapiens isoform 1 gene for L-type calcium channel, exon 47 and 48	0.069
578	S66283	Spnb-1=beta-spectrin [mice, reticulocyte, mRNA, 8126 nt]	0.069
579	M36305	Galago crassicaudatus gamma globin gene, complete cds.	0.07
580	AB018337.1	Homo sapiens mRNA for KIAA0794 protein, partial cds	0.22
581	Z60182	H.sapiens CpG island DNA genomic Mse1 fragment, clone 193a12, reverse read cpg193a12.rt1a	0.21
582	AF121948.1	Homo sapiens telomerase reverse transcriptase	0.003
584	Y10019	R.norvegicus mRNA for DRM protein	0.21
585	AF145653.1	Drosophila melanogaster clone GH08860 BcDNA.GH08860 (BcDNA.GH08860) mRNA, complete cds	0.64
586	M88321	Gossypium hirsutum group 4 late embryogenesis-abundant protein (Lea14-A) gene, complete cds.	0.024
587	S82821	GSTA5=glutathione S-transferase Yc2 subunit {5' region, intron 1} [rats, Morris hepatoma cell line, Genomic, 2212 nt, segment 1 of 3]	1.9
588	AF039857	5 Homo sapiens retinal pigment epithelium-specific protein (RPE65) gene, exon 3	0.023
589	X05034	Rat C2A gene for prostatic binding protein (PBP)	0.2
590	U13177	Rattus norvegicus clone ubc4a ubiquitin conjugating enzyme (E217kB) mRNA, complete cds.	0.071
591	AF081530	Homo sapiens neuralized binding protein mRNA, complete cds	e-143
592	Z73328	H.sapiens DNA (chromosome 13q, clone 117A11, 856 bp)	0.023
593	D49733	Mouse lamin A/C and C2 genes, exon 6, 7, 8, 9, 10, 11 and 12, complete cds	2.3
594	L04603	Trypanosoma cruzi R27-2 protein gene, complete cds.	2.3
595	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.72
596	D83993	Fission yeast DNA for chromosome II cosmid 1228 sequence	0.7

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
597	L77036	Homo sapiens (subclone 5_d9 from P1 H19) DNA sequence.	0.008
598	AE001414	Plasmodium falciparum chromosome 2, section 51 of 73 of the complete sequence	0.008
599	AF001893	Human MEN1 region clone epsilon/beta mRNA, 3' fragment	0.2
601	Z69652	Human DNA sequence from cosmid L75B9, Huntington's Disease Region, chromosome 4p16.3	0.023
602	Z16517	H. sapiens (D13S155) DNA segment containing	0.041
603	X14448	Human GLA gene for alpha-D-galactosidase A (EC 3.2.1.22)	0.71
604	AF055481	Homo sapiens normal epithelial cell-specific 1	0.029
605	AJ002550	Homo sapiens MMP-1 gene, promoter region	6E-11
606	AF037454	Mus musculus ubiquitin protein ligase (Itch) mRNA, complete cds	0.0009
607	U96108	Staphylococcus carnosus (3R)-hydroxymyristoyl acyl carrier protein dehydrase homolog (fabZ) gene, partial cds, YwpF homolog, single-strand binding protein homolog Sce...	0.8
608	X53334	Chicken mRNA for annexin II	0.029
609	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.028
610	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.085
611	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.089
612	D16474	Human mRNA, Xq terminal portion	0.00003
613	NM_004955.1	Homo sapiens equilibrative nucleoside transporter 1 (ENT1) mRNA >gi|1845344|gb|U81375|HSU81375 Human placental equilibrative nucleoside transporter 1	0.00003
614	AB019944.1	Arabidopsis thaliana gene for sigma factor SigC, complete cds	1.9
615	AB012181	Homo sapiens DNA, anonymous heat-stable fragment RP8-6A	1E-34
616	AF106929.1	Medicago truncatula putative cell wall protein (AM1) mRNA, complete cds	0.2
617	L09105	Homo sapiens glucos phosphate isomerase mRNA, intron with a conserved tandem repeat.	0.00003
618	X06292	Human c-fes/fps proto-oncogene	0.028
619	NM_003951.1	Homo sapiens solute carrier family 25 member 14 (SLC25A14), nuclear gene encoding mitochondrial product, mRNA mitochondrial carrier protein-1 (BMCP1) mRNA, nuclear gene encoding mitochondrial protein, complete cds	e-173
620	AL050089.1	Homo sapiens mRNA; cDNA DKFZp586E0518 (from clone DKFZp586E0518)	e-166
621	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.25
622	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.26
623	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.009

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
624	D32056	Human gene for 2-oxoglutarate dehydrogenase, exon 1 sequence	0.003
625	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.027
626	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.028
627	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.028
628	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.029
629	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.028
630	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.028
631	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.029
632	Z69364	Human DNA sequence from cosmid L96F8, Huntington's Disease Region, chromosome 4p16.3 contains EST and cDNA >gi|1182000|emb|Z69365|HSL96F8A Human DNA sequence from cosmid L96F8, Huntington's Disease Region, chromosome 4p16.3 contains EST and cDNA	8E-13
633	NM_004435.1	Homo sapiens endonuclease G (ENDOG), nuclear gene encoding mitochondrial protein, mRNA G (ENDOG) mRNA	9E-13
634	U46837	Human RNA polymerase II holoenzyme component SRB7 (SRB7) mRNA, complete cds.	0.21
635	M13973	Bovine protein kinase C mRNA, complete cds.	3E-14
636	AB012917	Homo sapiens mRNA for serine protease (TLSP), complete cds	e-143
637	M57750	S.pombe cut2+ gene, complete cds.	0.22
638	V00584	Human gene hY1 encoding a cytoplasmic Ro RNA	7E-21
639	L81854	Homo sapiens (subclone 2_b8 from P1 H48) DNA sequence	2E-11
640	X73897	H.sapiens zinc finger domain ZF21.3 DNA	2E-31
641	L10239	Insertion sequence IS1141 (from Mycobacterium intracellulare strain Va14), transposase gene, complete cds, clone pVT365.	1.8
642	AF097025	Homo sapiens cysteine desulfurase (nifS) mRNA, complete cds	e-170
644	AF008219	Borrelia afzelii R-IP3 chromosome right end, arcA and arcB genes, complete cds	0.092
645	NM_003496.1	Homo sapiens Transformation/transcription domain-associated protein (TRRAP) mRNA, and translated products >gi|4165076|gb|AF076974|AF076974 Homo sapiens TRRAP protein (TRRAP) mRNA, complete cds	6E-43
646	AF000305.1	Brassica napus steroid sulfotransferase 1 gene, complete cds	0.76
647	AF016031	Homo sapiens thyroid hormone receptor activator molecule (TRAM-1) mRNA, complete cds	8E-34
648	M97168	Homo sapiens X (inactive)-specific transcript	0.22

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
649	NM_003011.1	Homo sapiens SET translocation (myeloid leukemia-associated) (SET) mRNA cds.	9E-36
650	NM_004669.1	Homo sapiens chloride intracellular channel 3 (CLIC3) mRNA >gi|4323621|gb|AF102166|AF102166 Homo sapiens intracellular chloride channel CLIC3 (CLIC3) mRNA, complete cds	4E-50
651	AJ010479.1	Homo sapiens mRNA for kinesin-like protein 2	e-171
652	U29932	Human AMP deaminase (AMPD3) gene, intron 2, partial sequence.	1E-37
653	AF028233	Homo sapiens distal-less homeobox protein (DLX3) gene, complete cds	3E-47
654	AF151978.1	Homo sapiens amino acid transporter B0+	e-165
655	Z64037	H.sapiens CpG island DNA genomic Mse1 fragment, clone 95g8, forward read cpg95g8.ft1a	2E-50
656	M32140	T.brucei heat shock protein (Hsp70) gene, upstream region.	1.9
657	NM_003164.1	Homo sapiens syntaxin 5A (STX5A) mRNA mRNA, complete cds	7E-54
658	NM_001538.1	Homo sapiens heat shock transcription factor 4 (HSF4) mRNA >gi|1813425|dbj|D87673|D87673 Homo sapiens mRNA for heat shock transcription factor 4, complete cds	1E-57
659	NM_001538.1	Homo sapiens heat shock transcription factor 4 (HSF4) mRNA >gi|1813425|dbj|D87673|D87673 Homo sapiens mRNA for heat shock transcription factor 4, complete cds	1E-57
660	L76569	Homo sapiens (clones cYG3, B5P6C4) fragile X E mental retardation syndrome protein (FMR2) mRNA, complete cds.	0.21
661	X55110	Human mRNA for neurite outgrowth-promoting protein	2E-59
662	L20468	Rattus norvegicus cerebroglycan mRNA, complete cds.	3E-86
663	NM_005324.1	Homo sapiens H3 histone, family 3B (H3.3B)	e-127
664	NM_001283.1	Homo sapiens clathrin-associated/assembly/adaptor protein, small 1 Homo sapiens mRNA for sigma1A subunit of AP-1 clathrin adaptor complex, complete cds	e-171
665	AF007867	Lymantria dispar pheromone binding protein 1	1.8
669	U93704	Riftia pachyptila endosymbiont bacterioferritin comigratory protein homolog (bcp), sensor protein RssA complete cds	1.9
670	AB002315	Human mRNA for KIAA0317 gene, complete cds	1.8
673	X96585	M.musculus mRNA for NOV protein	1.8
674	D84103	Homo sapiens mRNA for mitochondrial DNA polymerase gamma, complete cds	1.7

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
675	AB000834.2	Nicotiana tabacum gene for thaumatin-like protein SE39b, complete cds	1.8
676	AF129853.1	Gymnascella hyalinospora strain VAMH 7366 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gen...	0.2
678	AB029007.1	Homo sapiens mRNA for KIAA1084 protein, complete cds	e-168
679	AB007957	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0488	e-145
680	AL080168.1	Homo sapiens mRNA; cDNA DKFZp434C151 (from clone DKFZp434C151)	0
681	D32166.1	Poplar mRNA for cellulase (endo-1, 4-beta-glucanase), complete cds	1.6
683	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.03
684	U32792	Haemophilus influenzae Rd section 107 of 163 of the complete genome	2.1
685	X74969	R.norvegicus gene for prostatic acid phosphatase	0.02
686	U70998	Phanerochaete chrysosporium manganese peroxidase isozyme 3 (mnp3) gene, complete cds	0.73
687	NM_005969.1	Homo sapiens nucleosome assembly protein 1-like 4 (NAP1L4) mRNA >gi|1679778|gb|U77456|HSU77456 Human nucleosome assembly protein 2 mRNA, complete cds	2.1
688	AJ132369.1	Sorites orbiculus SSU rRNA, isolate 206	0.67
690	U04435	Drosophila melanogaster GLI-Kr zinc finger pair-rule protein mRNA, complete cds. embryo, mRNA, 2959 nt]	0.67
691	X69511	G.gallus Acra-2 gene alpha-2 subunit	0.67
692	AF140762.1	Homo sapiens neuronal acetylcholine receptor beta-3 subunit precursor (CHRNB3) gene, exon 3	2
695	Z74734	C.porcellus mRNA for guanylyl cyclase C	1.9
696	L76081	Clostridium difficile ADP-ribosyltransferase enzymatic and binding component (cdtA and cdtB) genes, complete cds's	0.63
697	X82657	H.sapiens IRLB gene (exon 4)	0.66
698	AB020649.1	Homo sapiens mRNA for KIAA0842 protein, partial cds	e-143
699	NM_005499.1	Homo sapiens SUMO-1 activating enzyme subunit 2 (UBA2) mRNA >gi|4096671|gb|U35832.1|HSU35832 Human anthracycline-associated resistance ARX mRNA, complete cds	1E-47

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
700	AB018255.1	Homo sapiens mRNA for KIAA0712 protein, complete cds	0.008
701	AL035496.6	Human DNA sequence from clone 437O22 on chromosome 22q12.2-13.1. Contains the 5' part of a novel VHS domain containing protein similar to predicted worm and human proteins. Contains ESTs, GSSs and a putative CpG islan...	0.0000001
702	AB020664.1	Homo sapiens mRNA for KIAA0857 protein, partial cds	e-162
703	AL050269.1	Homo sapiens mRNA; cDNA DKFZp564C103 (from clone DKFZp564C103)	e-173
704	M62324	Human modulator recognition factor I (MRF-1) mRNA, 3' end.	1.8
705	Z69363	Human DNA sequence from cosmid L60G9B, Huntington's Disease Region, chromosome 4p16.3 contains ESTs	0.61
707	AF068890	Bos taurus PIM1 protein (PIM1) gene, exon 5 and partial cds	0.64
708	NM_000211.1	Homo sapiens integrin beta chain, beta 2 leukocyte adhesion protein (LFA-1/Mac-1/p150,95 family) beta subunit mRNA.	0.65
709	U38550	Arabidopsis thaliana pre zeta-carotene desaturase precursor (zds) mRNA, complete cds.	1.9
710	AF147787.1	Homo sapiens hepatocyte nuclear factor-3 beta gene, complete cds	0.22
711	AF140549.1	Enterococcus faecium unknown gene	0.19
712	AF031630	Danio rerio homeobox protein LIM-3 (lim3) gene, exons 2 and 3	0.19
713	AF007883	Homo sapiens MHC class II HLA-DRB1 (HLA-DRB1*10) intron 1 sequence	0.021
714	X71844	C.perfringens uapC, cpe, and nadC genes	0.63
715	M87359	Yeast Eco RI fragment.	0.56
716	AF088887	Oryctolagus cuniculus interleukin-10 precursor, mRNA, complete cds	0.62
717	AF151897.1	Homo sapiens CGI-139 protein mRNA, complete cds	3E-38
718	U65948	Zea mays starch branching enzyme IIa (Sbe2a) mRNA, partial cds	0.61
719	AF086443	Homo sapiens full length insert cDNA clone ZD81C11	1E-68
720	AJ011767	Sus scrofa mRNA for neuron-derived orphan receptor-1 alfa transcription factor	0.18
721	U78547	Chlamydomonas reinhardtii PF20 mRNA, complete cds	0.00009
722	U25686	Drosophila melanogaster ecdysone-regulated (E93) mRNA, complete cds.	0.54

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
723	AB029017.1	Homo sapiens mRNA for KIAA1094 protein, complete cds	e-102
724	U67533	Methanococcus jannaschii section 75 of 150 of the complete genome	0.4
725	L81892	Homo sapiens (subclone 2_h6 from P1 H62) DNA sequence	2.2
726	U83650	Mus caroli Sp100 gene, exon 13	2.1
727	X14710	B.taurus beta-lactoglobulin gene	0.23
728	AF034920	Homo sapiens tubby like protein 1 (TULP1) gene, exons 9-11	2
729	D83999	Mus musculus mRNA for the third largest RNA polymerase II subunit, complete cds	0.22
730	U18109	Macropus rufogriseus MHC class II DR alpha protein precursor (Maru-DRA) mRNA, complete cds.	0.66
731	Y18476	Trichophyton rubrum mitochondrial cytb gene and NADH1 to NADH5 genes	0.67
732	Y18476	Trichophyton rubrum mitochondrial cytb gene and NADH1 to NADH5 genes	0.65
733	AF148461.1	Homo sapiens CLNS1A gene, intron 1 sequence	e-160
734	L80007	Equine adenovirus 2 385/75 hexon and endopeptidase genes, complete cds	1.9
735	X76128	T.thermophila MSE 2.9 (left) gene germline limited sequence	0.22
737	AF141658.1	Ictalurus punctatus EB1 mRNA, complete cds	0.62
738	U67576	Methanococcus jannaschii section 118 of 150 of the complete genome	0.21
739	AL050269.1	Homo sapiens mRNA; cDNA DKFZp564C103 (from clone DKFZp564C103)	e-159
740	AF065389	Homo sapiens tetraspan NET-4 mRNA, complete cds	0.21
741	AB007455.1	Homo sapiens mRNA for P53TG1-A, complete cds	0.22
742	U67399	Mus musculus K-cadherin/cadherin-6 mRNA, partial cds	2
743	AB018315.1	Homo sapiens mRNA for KIAA0772 protein, complete cds	9E-78
745	AL080164.1	Homo sapiens mRNA; cDNA DKFZp564C1940 (from clone DKFZp564C1940)	5E-20
746	X00007	Bacillus subtilis 5' end of ribosomal RNA operon rrnB	0.22
747	AF058234.1	Scutellastra longicosta 16S ribosomal RNA gene, mitochondrial gene for mitochondrial RNA, partial sequence	0.022
748	M99362	Rhesus macaque polyoma virus large T antigen gene, 3' end.	0.2
749	U80458	Human microtubule associated protein 1A mRNA, partial cds	0.067

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
750	AB019533.1	Oryza sativa mRNA for Nad-dependent formate dehydrogenase, complete cds	0.22
751	Z69723	Human DNA sequence from cosmid U238E5, between markers DXS6791 and DXS8038 on chromosome X	0.2
752	AF056936	Plasmodium falciparum mature parasite-infected erythrocyte surface antigen gene, complete cds	1.8
753	AJ010396.1	Homo sapiens DKC1 gene, exons 12 to 15	0.63
754	U19253	Xenopus laevis/gilli complement component C3 mRNA, partial cds.	1.9
755	M82872	S.cerevisiae protein-tyrosine phosphatase complete cds.	0.21
756	AF045188	Salmo salar ribosomal protein L18a mRNA, complete cds	0.21
757	AJ001118	Mus musculus mRNA for monoglyceride lipase	0.62
758	Y10377	C.albicans TOP2 gene	1.8
759	AB014573	Homo sapiens mRNA for KIAA0673 protein, partial cds	e-168
760	L24113	Saccharomyces cerevisiae Ca2+ regulatory protein	0.19
761	M96739	Human NSCL-1 mRNA sequence.	1.7
762	AF035006	Human respiratory syncytial virus, recombinant mutant rA2cp, complete genome	0.56
763	AF065389	Homo sapiens tetraspan NET-4 mRNA, complete cds	0.19
764	NM_006354.1	Homo sapiens transcriptional adaptor 2 complex) (TADA3L) mRNA >gi|3335554|gb|AF069733|AF069733 Homo sapiens ADA3-like protein mRNA, complete cds	e-154
765	M73752	Gossypium hirsutum Lea4-A gene, complete CDS.	0.06
766	X79192	F.brownii pdk gene	0.54
767	X14891	H.sapiens gene for transforming growth factor-beta 3 (TGF-beta 3) exon 7	0.076
768	Z78708	H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA14H12	0.076
769	AF068902	Streptococcus pneumoniae D-glutamic acid adding enzyme MurD (murD), undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase (murG), cell division protein DivIB (divIB), orotidine-5'-decarboxylase PyrF (pyrF), an...	0.23
770	U62588	Cricetulus griseus beta-1,6-N-acetylglucosaminyltransferase Lec4 cell line insertion mutant mRNA, complete cds	2
771	AJ236354.1	Timarcha coarcticollis mitochondrial partial tRNA-Leu gene and COII gene, isolate Los Barrios, Cadiz, Spain	0.026
772	U67604	Methanococcus jannaschii section 146 of 150 of the complete genome	0.22

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
773	AB018257.1	Homo sapiens mRNA for KIAA0714 protein, partial cds	e-178
774	AF169299.1	Equus caballus microsatellite HTG15 sequence	0.21
775	U96289	Homo sapiens Ig heavy chain VH3 region (VH3-30.3) mRNA, partial cds	0.64
776	Y07521	Mouse neuroblastoma-Rat glioma hybrid cell line mRNA for a potassium channel protein NGK2	0.076
777	NM_003966.1	Homo sapiens sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain semaphorin F homolog mRNA, complete cds	0.071
778	U66524	Dictyostelium discoideum ORFveg158 mRNA, partial cds	0.071
779	M81388	Chilo iridescent virus DNA-directed RNA polymerase and helicase genes, complete cds's. DNA-depenent RNA polymerase largest subunit homolog iridescent virus type 6, Genomic, 3 genes, 7990 nt]	0.073
780	AF132944.1	Homo sapiens CGI-10 protein mRNA, complete cds	e-170
781	AJ001700	Mus musculus mRNA for neuroserpin	0.069
782	U50421	Human Down Syndrome region of chromosome 21, clone A4B8-1D8.	0.61
783	X74159	K.lactis MBP1 gene	1.9
784	D87682	Human mRNA for KIAA0241 gene, partial cds	0.071
785	AB015633.1	Homo sapiens mRNA for type II membrane protein, complete cds, clone:HP10481	7E-23
786	AF022414	Trichomonas vaginalis glyceraldehyde-3-phosphate dehydrogenase (gap2) gene, partial cds	0.2
787	M36996	Mouse L1M1 and L1M2 sequence DNA.	0.21
788	M95171	Aedes aegypti LINE retrotransposon Juan-A including DNA binding protein and reverse transcriptase-like protein mRNA, complete coding regions.	0.069
789	U47661	Lupinus luteus proline-rich protein PRP2 precursor (LlPRP2) gene, complete cds	0.59
790	X55581	H.sapiens immunoglobulin heavy chain gene, diversity region	0.59
791	AF104500	Farfantepenaeus duorarum isolate FD6 mitochondrial control region	0.065
792	AF094519	Mus musculus diaphanous-related formin (Dia2) mRNA, complete cds	3E-79
793	X95267	G.gallus mRNA for ryanodine receptor type 3	0.63
794	K01872	Bacteriophage Cp-1 (Streptococcus pneumoniae), 3' inverted terminal repeat.	0.063

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
795	AF094573	Rice tungro bacilliform virus isolate T10 P194 gene, partial cds	1.7
796	X15061	Glycine max lbc3 gene for leghemoglobin C3	0.062
797	Y17267	Mus musculus mRNA for ubiquitin conjugating enzyme	3E-89
798	AF149109.1	Rickettsia australis strain PHS outer membrane protein B (ompB) gene, partial cds	0.061
799	AJ004870	Thermoanaerobacterium thermosaccharolyticum ptaA and ackA genes, orf1, orf2, orf3, orf4	0.19
800	S77555	corticotropin receptor/ACTH receptor {5' region}	0.19
801	AJ130796	Mus musculus APC2 gene, exon 14	1.6
802	AE001229	Treponema pallidum section 45 of 87 of the complete genome	1.7
803	X16137	Suillus sinuspaulianus mitochondrial large subunit ribosomal RNA gene, part	0.69
804	AF048839.1	Arabidopsis thaliana Atmyb103 (MYB103) gene, complete cds	0.68
805	Z86109	S.carlsbergensis 12 kb region of chromosome III	0.025
806	X67053	S.tuberosum ppc mRNA for phosphoenolpyruvate carboxylase	0.67
807	X13423	Phaseolus vulgaris tRNA-Pro(UGG3) gene	2
808	Z17118	H. sapiens (D9S179) DNA segment containing (CA) repeat; clone AFM248wf1; single read	0.65
809	Z23386	H. sapiens (D5S467) DNA segment containing	0.072
810	M12729	Mouse T-cell surface antigen T3 delta-chain gene, exons 2,3,4 and 5, from B8C3 (anti-porcine insulin T-T) hybridoma, clone pMT-2.	0.23
811	AF012551	Plasmodium falciparum ornithine decarboxylase	0.21
812	L08265	Human skeletal muscle chloride channel (HUMCLC) gene, exon 7.	0.075
813	X00616	Tobacco chloroplast gene P32 for thylakoid membrane protein	0.07
814	U79731	Plasmodium berghei extrachromosomal plastid PB-1, ORF470 gene, partial cds, tRNA-Thr, large subunit ribosomal RNA, tRNA-Met, tRNA-Arg, tRNA-Val, tRNA-Arg, tRNA-Leu, tRNA Asn, tRNA-Ala, and small subunit ribosomal RNA genes...	0.023
815	S82293	II beta-globin=II beta-globin {5' region} [rats, mRNA Partial, 1428 nt]	2
816	AJ007398.1	Homo sapiens mRNA for PBK1 protein	0
817	AL109849.1	Streptomyces coelicolor cosmid 3A3	0.023

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
819	U84223	Canine herpesvirus cIR6, cUS2, cUS3, cUS4, cUS6, cUS7, cUS8 and cUS9 genes, complete cds	0.067
820	AF036233	Homo sapiens cdc25B phosphatase (CDC25B) gene, alternatively spliced, partial cds	0.024
821	AB016195.1	Homo sapiens ELK1 pseudogene (ELK2) and immunoglobulin heavy chain gamma pseudogene (IGHGP)	1E-16
822	L10820	Human N-formyl peptide receptor (FPR1) gene, complete cds and Alu repeats.	0.023
823	AF039423	Cebus olivaceus blue opsin gene, exons 2 and 3	0.58
824	Z97214	Xenopus laevis mRNA for MILZ protein	1.8
825	Y17038	Mus musculus bassoon gene, exon 6 to 11	1.7
826	X92518	H.sapiens mRNA for HMGI-C protein	0.065
827	X06414	Mycoplasma capricolum ribosomal protein gene cluster	0.62
828	AJ002019	Saccharomyces uvarum mitochondrial coxII gene, partial	0.061
829	D84395	Bombyx mori DNA for cecropin A, complete cds	0.18
830	D86077	Homo sapiens DNA for cyclin G, partial cds	0.18
831	AF144573.1	Mesocricetus auratus Mx-interacting protein kinase PKM mRNA, complete cds	2E-18
832	AF123653.1	Homo sapiens FEZ1 (FEZ1) gene, complete cds	0.009
833	X71018	N.tabacum NPG-G27Y mRNA for polygalacturonase	0.025
834	D63884	Anthocidaris crassispina mRNA for intermediate chain 1, complete cds	0.072
835	AE001393	Plasmodium falciparum chromosome 2, section 30 of 73 of the complete sequence	0.008
836	AE001395	Plasmodium falciparum chromosome 2, section 32 of 73 of the complete sequence	0.0003
837	AF007164	Drosophila melanogaster mRNA sequence	0.21
838	AB005744	Perilla frutescens DNA for 1-limonene synthase, complete cds	0.21
839	AF067143	Homo sapiens myosin heavy chain (MYH8) gene, partial cds	0.21
840	X04130	Watermelon mitochondrial URF1 gene	0.008
841	X95439	S.xylosus aroA, ccpA, acuC and acuA genes	0.008
842	AB007404.1	Oryza sativa gene for alanine aminotransferase, complete cds	0.063
843	X59823	Human chromosome 8 flanking hypervariable simple repeat DNA (clone HZREP32)	0.21
844	AF122981.1	Arabidopsis lyrata cultivar NC4 RPM1 gene, 5' sequence	0.002
845	AF016667.2	Caenorhabditis elegans cosmid T20H12	4.9

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P. VALUE
846	V00184	Slime mold (D. discoideum) gene for actin 2 sub1 actin 2 (sub 1) gene 5' end.	0.061
847	AB020656.1	Homo sapiens mRNA for KIAA0849 protein, partial cds	0.23
848	X82107	H.sapiens gene for tryptophanyl-tRNA synthetase	1.9
849	M81385	Mouse liver receptor homologous protein (LRH-1) mRNA, complete cds.	0.025
850	NM_006401.1	Homo sapiens acidic protein rich in leucines silver-stainable protein SSP29 mRNA, complete cds	0.008
851	AL010149	Plasmodium falciparum DNA *** SEQUENCING IN PROGRESS *** from contig 3-82, complete sequence	0.21
852	AF151826.1	Homo sapiens CGI-68 protein mRNA, complete cds	e-153
853	AE001402	Plasmodium falciparum chromosome 2, section 39 of 73 of the complete sequence	0.021
854	U96976	Homo sapiens MET proto-oncogene, intron 6, 3' end	0.068
855	D85545	Yeast chk1 and ucbP4 DNA, partial and complete cds	1.7
856	AF001175	Homo sapiens ribonuclease P protein subunit p14 (Rpp14) mRNA, complete cds	7E-45
857	U14724.1	Anticarsia gemmatalis nuclear polyhedrosis virus genomic repeat region	0.0008
858	AL008641	Human DNA sequence from cosmid N100B10 on chromosome 22q12.3	0.06
859	Y00326	Human sis proto-oncogene upstream region	0.19
860	AF003483	Habrabracon hebetor 16S ribosomal RNA gene, partial sequence	0.0007
861	AL049265.1	Homo sapiens mRNA; cDNA DKFZp564F053 (from clone DKFZp564F053)	e-122
862	Z69351	B.vulgaris repetitive DNA (clone pDRV1)	0.0009
863	L28998	Theileria parva 28S ribosomal RNA (28S rRNA) gene.	0.024
864	X06000	G.gallus carbonic anhydrase II gene exons 1-2	0.067
865	AB000565	Homo sapiens DNA for repeat sequence Alu	1E-26
866	AE000761	Aquifex aeolicus section 93 of 109 of the complete genome	0.22
867	AF017145	Homo sapiens multidrug resistance protein	0.0008
868	J05451	Human gastric (H+ + K+)-ATPase gene, complete cds.	0.003
869	AB018258.1	Homo sapiens mRNA for KIAA0715 protein, partial cds	0.007
870	D78572	Mus musculus mRNA for membrane glycoprotein, complete cds >gi|3251779|dbj|E12950|E12950 cDNA GA3-43 encoding novel polypeptide which appear when differentiate from embryo-tumor cell P19 to nerve cell	0.0001

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
871	AC001017	Homo sapiens (subclone 2_g8 from P1 H43) DNA sequence	0.003
872	AB018284.1	Homo sapiens mRNA for KIAA0741 protein, complete cds	0.0009
873	Z54147	Human DNA sequence from cosmid L129H7, Huntington's Disease Region, chromosome 4p16.3 contains CpG island	0.002
874	Z54147	Human DNA sequence from cosmid L129H7, Huntington's Disease Region, chromosome 4p16.3 contains CpG island	0.002
875	NM_006392.1	Homo sapiens nucleolar protein (KKE/D repeat) mRNA for nucleolar protein hNop56	e-157
876	L43392	Homo sapiens (subclone 6_a8 from P1 H16) DNA sequence.	0.00001
877	X75670	O.sativa mRNA for cytochrome b5	0.00001
878	Z23808	H. sapiens (DXS1199) DNA segment containing	0.000009
879	L48473	Homo sapiens (subclone 7_e11 from P1 H16) DNA sequence.	0.003
880	AF074908.1	Homo sapiens neuronal and epithelial glutamate transporter (SLC1A1) gene, exon 7	5E-11
881	X02536	Human preproenkephalin B gene 5' region and exon 1 >gi|182100|lcl|X00174 Human enkephalin B (enkB) gene, 5' flank and exon 1.	0.000001
882	AL109681.1	Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 112333	0.000003
883	M88599	Entamoeba histolytica P-glycoprotein-1 (pgp1) gene, complete cds.	0.07
884	M38188	Human unknown protein from clone pHGR74 mRNA, complete cds.	6E-10
885	U50531	Human BRCA2 region, mRNA sequence CG030	0.000001
886	Z64533	H.sapiens CpG island DNA genomic Mse1 fragment, clone 134d9, forward read cpg134d9.ft1a	0.0000004
887	NM_004422.1	Homo sapiens dishevelled 2 (homologous to Drosophila dsh) (DVL2) mRNA dishevelled 2 (DVL2) mRNA, complete cds	e-116
888	S66168	sterol regulatory element 1 binding protein cells, mRNA Partial, 547 nt, segment 2 of 2] 5527690	0.008
889	L29556	Human (clone hSTX) sialyltransferase mRNA, 3' end.	0.008
890	AF036703	Caenorhabditis elegans cosmid T11F8	0.53
891	Z68281	Human DNA sequence from cosmid L2F10, Huntington's Disease Region, chromosome 4p16.3 contains Human G protein coupled receptor kinase-like, and an RFLP	0.000004
892	AL035046.5	Human DNA sequence from clone 321I20 on chromosome 1q32.1-41 Contains GSSs, complete sequence	0.0001
893	Y15083	Homo sapiens p14.5-like gene and Alu repeat	3E-13

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
894	AF044123	Homo sapiens clone SUPTH48 sequence flanking the HIV-1 provirus integration site	0.19
895	J00139	Human dihydrofolate reductase gene, exon 6 and 3' flank.	7E-44
896	Y16790	Homo sapiens hHa4 gene, complete CDS	3E-14
897	U50105	Human ankyrin (ANK1) gene, exon 15	0.00000004
898	AB000537	Schizosaccharomyces pombe mRNA for snoRNP protein GAR 1, complete cds	0.00000005
899	D79990	Human mRNA for KIAA0168 gene, complete cds	0.000001
900	AC002183	Homo sapiens (subclone 2_h8 from BAC H111) DNA sequence	0.00000004
901	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.11
902	AF072468	Homo sapiens (JH8) mRNA, partial cds	2E-19
903	M11167	Human 28S ribosomal RNA gene.	2E-09
904	D87117	Mus musculus mRNA for SAP102, complete cds	2E-09
905	AB023189.1	Homo sapiens mRNA for KIAA0972 protein, complete cds	0
906	Y07554	Psychrobacter sp. pim gene	0.68
907	AC005190	Homo sapiens PAC clone DJ1152D16 from Xq23, complete sequence [Homo sapiens]	2E-29
908	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.28
909	AF135183.1	Homo sapiens Recq helicase 5 (RECQ5) gene, alternative splice products, complete cds	e-146
910	U58884	Mus musculus SH3-containing protein SH3P7 mRNA, complete cds. similar to Human Drebrin	2E-13
911	U29113	Human leiomyoma cell line LM-30.1/SV40 ectopic sequence from HMGI-C fusion mRNA, 3' sequence, clone pCH110.	2E-13
912	AC002252	Homo sapiens (subclone 1_g7 from BAC H76) DNA sequence	3E-24
913	U95097	Xenopus laevis mitotic phosphoprotein 43 mRNA, partial cds	0.09
914	NM_003437.1	Homo sapiens zinc finger protein 136 (clone pHZ-20) (ZNF136) mRNA >gi|487784|gb|U09367|HSU09367 Human zinc finger protein ZNF136	2E-19
915	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.029
916	M12523	Human serum albumin (ALB) gene, complete cds.	1E-15
917	AF043324	Homo sapiens N-myristoyltransferase 1 mRNA, complete cds	2E-51
918	L43631	Homo sapiens scaffold attachment factor B (SAF-B) mRNA, partial cds	0.008
919	L01616	Tribolium castaneum zinc finger protein (Kruppel domain region) gene, partial cds.	4E-18
920	D12688	Mouse P-cadherin gene, exon 1 and 2	2
921	X94770	H.sapiens mRNA for epithelial membrane protein-2	2E-19

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
922	NM_000946.1	Homo sapiens primase, polypeptide 1 (49kD) for DNA primase (subunit p48)	2E-20
923	AE000818	Methanobacterium thermoautotrophicum from bases 264585 to 276866 (section 24 of 148) of the complete genome	1.9
924	AL035418.6	Human DNA sequence from clone 141I3 on chromosome 22q13.1-13.33 Contains an STS and a GSS, complete sequence [Homo sapiens]	0.0009
925	AF086040	Homo sapiens full length insert cDNA clone YX52E07	6E-73
926	AL049310.1	Homo sapiens mRNA; cDNA DKFZp564B206 (from clone DKFZp564B206)	2E-09
927	J01415	Human mitochondrion, complete genome	3E-24
928	AF044122	Homo sapiens clone SUPTH47 sequence flanking the HIV-1 provirus integration site	9E-25
929	Z22640	H.magnipapillata homeobox containing exon	0.076
930	NM_004859.1	Homo sapiens clathrin, heavy polypeptide-like 2 (CLTCL2) mRNA >gi|434760|dbj|D21260|HUMORFEA Human mRNA for KIAA0034 gene, complete cds	4E-27
931	AL049701.1	Human gene from PAC 433G19, chromosome 1	e-162
932	NM_004698.1	Homo sapiens U4/U6-associated RNA splicing factor (HPRP3P) mRNA >gi|2708306|gb|AF016370|AF016370 Homo sapiens U4/U6 small nuclear ribonucleoprotein hPrp3 mRNA, complete cds	4E-28
933	NM_003983.1	Homo sapiens solute carrier family 7 for KIAA0245 gene, complete cds	1E-30
934	NM_003440.1	Homo sapiens zinc finger protein 140 (clone pHZ-39) (ZNF140) mRNA >gi|487786|gb|U09368|HSU09368 Human zinc finger protein ZNF140	1E-30
935	AL050392.1	Homo sapiens mRNA; cDNA DKFZp586I031 (from clone DKFZp586I031)	7E-33
936	NM_002714.1	Homo sapiens protein phosphatase 1, regulatory subunit 10 (PPP1R10) mRNA	e-121
937	M27830	Human 28S ribosomal RNA gene, complete cds.	2E-33
938	M27830	Human 28S ribosomal RNA gene, complete cds.	2E-33
939	Z72521	Human DNA sequence from cosmid N29F4 on chromosome 22q11.2-qter contains STS	0.000001
940	AF056195	Homo sapiens neuroblastoma-amplified protein mRNA, complete cds	2E-72
941	Z35989	S.cerevisiae chromosome II reading frame ORF YBR120c	0.19
942	U76557	Rattus norvegicus O-GlcNAc transferase, p110 subunit (OGT) mRNA, complete cds	9E-36

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
943	Z56141	H.sapiens CpG island DNA genomic Mse1 fragment, clone 8g7, forward read cpg8g7.ft1a	3E-37
944	AF132951.1	Homo sapiens CGI-17 protein mRNA, complete cds	e-165
945	NM_006548.1	Homo sapiens IGF-II mRNA-binding protein 2 sapiens hepatocellular carcinoma autoantigen (p62) mRNA, complete cds	e-140
946	AF010317	Homo sapiens Pig3 (PIG3) gene, partial cds	3E-38
947	AB023151.1	Homo sapiens mRNA for KIAA0934 protein, partial cds	2E-54
948	Z69649	Human DNA sequence from cosmid L69F7B, Huntington's Disease Region, chromosome 4p16.3 contains Huntington Disease (HD) gene	1E-25
949	AC001159	Homo sapiens (subclone 1_h9 from PAC H92) DNA sequence	4E-17
950	AL080060.1	Homo sapiens mRNA; cDNA DKFZp564H172 (from clone DKFZp564H172)	5E-29
951	L07758	Human IEF SSP 9502 mRNA, complete cds.	4E-48
952	M29037	Human 17 beta-hydroxysteroid dehydrogenase	0.56
953	M36704	C.perfringens perfringolysin O (pfo) gene, complete cds.	0.22
954	U34991	Human endogenous retrovirus clone c18.4, HERV-H/HERV-E hybrid multiply spliced protease/integrase mRNA, complete cds, and envelope protein mRNA, partial cds	2E-61
955	AB002369	Human mRNA for KIAA0371 gene, complete cds	0.0009
956	AF098668	Homo sapiens acyl-protein thioesterase mRNA, complete cds	e-156
957	L12019	Actinidia deliciosa var delicisoa polygalacturonase gene, complete cds	0.19
958	Z12622	A.sativum mRNA encoding precursor alliinase	0.065
959	NM_003429.1	Homo sapiens zinc finger protein 85 (HPF4, HTF1) (ZNF85) mRNA >gi|1017721|gb|U35376|HSU35376 Human repressor transcriptional factor (ZNF85) mRNA, complete cds.	2E-51
960	AP000249.1	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region, clone:B762O15, complete sequence	0.0003
961	U16120	Human placental taurine transporter mRNA, complete cds.	2E-52
962	NM_002286.1	Homo sapiens lymphocyte-activation gene 3 mRNA for CD4-related protein involved in lymphocyte activation	2E-53
963	D63876	Human mRNA for KIAA0154 gene, partial cds	6E-54
964	NM_004128.1	Homo sapiens general transcription factor IIF, polypeptide 2 (30kD subunit) (GTF2F2) mRNA subunit of transcription initiation factor RAP30/74	7E-55

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
965	AF004691	Scutellospora heterogama 18S ribosomal RNA gene, partial sequence, 5.8S ribosomal RNA gene, complete sequence, and 26S ribosomal RNA gene, partial sequence	0.22
966	U67509	Methanococcus jannaschii section 51 of 150 of the complete genome	0.074
967	AB010059	Homo sapiens RBP56/hTAFII68 gene, exon 3, 4, 5	4E-80
968	AB007891	Homo sapiens KIAA0431 mRNA, partial cds	9E-60
969	NM_005873.1	Homo sapiens G alpha interacting protein (GAIP) mRNA >gi|1107697|emb|X91809|HSPAIP H.sapiens mRNA for GAIP protein	4E-60
970	M76558	Human neuronal DHP-sensitive, voltage-dependent, calcium channel alpha-1D subunit mRNA, complete cds.	0.27
971	NM_003431.1	Homo sapiens zinc finger protein 124 (HZF-16) HZF-16=Kruppel-related zinc finger gene homolog HEP-G2, mRNA, 2080 nt]	2E-60
972	AB002584	Rattus norvegicus mRNA for beta-alanine-pyruvate aminotransferase, complete cds	0.00000002
973	X85133	H.sapiens RBQ-1 mRNA	2E-64
974	AJ130872.1	Porphyromonas gingivalis W50 receptor antigen (rag) locus encoding a major immunodominant 55kDa antigen	1.7
975	U67203	Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds	2E-66
976	U67203	Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds	3E-69
977	U55941	Expression vector pVP-HA2, complete sequence.	2E-79
978	AF074331.1	Homo sapiens PAPS synthetase-2 (PAPSS2) mRNA, complete cds	e-173
979	X78684	M.musculus mRNA for B-cell receptor associated protein (BAP) 29	e-100
980	AF060539	Mus musculus channel interacting PDZ domain protein mRNA, complete cds	e-138
981	U55042	Bos taurus myosin X, complete cds	e-119
982	AB000172	Porcine mRNA for endopeptidase 24.16, complete cds	e-131
983	Z57139	H.sapiens CpG island DNA genomic Mse1 fragment, clone 165d10, forward read cpg165d10.ft1a	0.4
984	AB012917	Homo sapiens mRNA for serine protease (TLSP), complete cds	0
985	L39064	Homo sapiens interleukin 9 receptor precursor	6E-15
986	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.36
987	AF045188	Salmo salar ribosomal protein L18a mRNA, complete cds	0.38
988	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.43

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
989	AF055287	Emericella nidulans molybdenum cofactor biosynthetic protein (cnxF) gene, complete cds	4.4
990	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.38
991	X03991	Human glucagon gene	0.42
992	AJ131021.1	Mus musculus mRNA for pp90 ribosomal protein S6 kinase 3	2E-17
993	U18168	Human HLA class I genomic survey sequence, contains Alu.	4E-11
994	Z57139	H.sapiens CpG island DNA genomic Mse1 fragment, clone 165d10, forward read cpg165d10.ft1a	0.4
995	AB012917	Homo sapiens mRNA for serine protease (TLSP), complete cds	0
996	AF045742	Xenopus laevis Smad7 mRNA, complete cds	0.38
997	M72411	Human MHC class II HLA-DQA1 gene (DR4,DR4), flanking region and alu repeat.	4E-21
998	J03612	P.yoelii merozoite surface antigen gene, 3' end.	0.13
999	AB007957	Homo sapiens mRNA, chromosome 1 specific transcript KIAA0488	0
1000	AF034265	Gracilaria chilensis 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 25S ribosomal RNA gene, partial sequence	0.62
1001	AL080168.1	Homo sapiens mRNA; cDNA DKFZp434C151 (from clone DKFZp434C151)	0
1002	AF121948.1	Homo sapiens telomerase reverse transcriptase	0.001
1003	AF063668.1	Mus musculus type XIII collagen (col13a1) gene, exon 3	2.1
1004	NM_003630.1	Homo sapiens peroxisomal biogenesis factor 3 mRNA for Pex3 protein	0
1005	M35543	Human GTP-binding protein (G25K) mRNA, complete cds.	0.077
1006	U68216	Carica papaya ACC synthase mRNA, complete cds	4
1007	AF133913.1	Mus musculus ARL-6 interacting protein-6	6E-82
1008	NM_000211.1	Homo sapiens integrin beta chain, beta 2 leukocyte adhesion protein (LFA-1/Mac-1/p150,95 family) beta subunit mRNA.	0.65
1009	D10044	Tomato aspermy virus (V-TAV) RNA1	0.02
1010	S82740	NPM/ALK=fusion gene {translocation breakpoint}	0.71
1011	U78776	Treponema denticola gufa gene, partial cds, putative flagellar operon flgB, flgC, fliE, fliF, fliG, fliH, fliI and fliJ genes, complete cds, and fdgA gene, partial cds	0.14

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1012	AF140549.1	Enterococcus faecium unknown gene	0.51
1013	AF124490.1	Homo sapiens ARF GTPase-activating protein GIT1 mRNA, complete cds	e-176
1014	D10044	Tomato aspermy virus (V-TAV) RNA1	0.02
1015	AF088887	Oryctolagus cuniculus interleukin-10 precursor, mRNA, complete cds	0.62
1016	AF121948.1	Homo sapiens telomerase reverse transcriptase	0.29
1018	U78547	Chlamydomonas reinhardtii PF20 mRNA, complete cds	0.0001
1019	AL080145.1	Homo sapiens mRNA; cDNA DKFZp434P113 (from clone DKFZp434P113)	0
1020	M26198	Bovine ASS mRNA encoding argininosuccinate synthetase, complete cds.	0.24
1021	D87686.1	Homo sapiens mRNA for KIAA0017 protein, complete cds	e-175
1022	X56668	Human DNA for calretinin exon 1	0.16
1023	AB007158	Homo sapiens gene for ribosomal protein S23, partial cds	e-114
1024	X83433	O.sativa mRNA for lipid transfer protein, b21	0.66
1025	D32072	Mouse mRNA for an isoform of TGF-b type II receptor	0.074
1026	D26077	Mouse mRNA for KIF3B protein, complete cds	0.3
1027	X80111	D.melanogaster sap47-1 mRNA	2E-09
1028	NM_003951.1	Homo sapiens solute carrier family 25 member 14 (SLC25A14), nuclear gene encoding mitochondrial product, mRNA >gi|3851539|gb|AF078544|AF078544 Homo sapiens brain mitochondrial carrier protein-1 (BMCP1) mRNA, nuclear gene encoding mitochondrial protein, complete cds	0
1029	AF016422	Caenorhabditis elegans cosmid R09E12	0.0007
1030	NM_006354.1	Homo sapiens transcriptional adaptor 2 (ADA2, yeast homolog)-3 like (PCAF histone acetylase complex) sapiens ADA3-like protein mRNA, complete cds	0
1031	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.063
1032	M25702	Human thyroid peroxidase (TPO) gene, exon 2.	0.091
1033	AB018257.1	Homo sapiens mRNA for KIAA0714 protein, partial cds	0
1034	L06898	Actinomyces viscosus sialidase (nanH) gene, complete cds.	0.49
1035	Y07521	Mouse neuroblastoma-Rat glioma hybrid cell line mRNA for a potassium channel protein NGK2	0.12
1036	AF153201.1	Homo sapiens zinc finger protein dp mRNA, complete cds	3E-39
1037	AJ010642	Drosophila melanogaster mRNA for Dof protein, transcript I, partial	1.9

Table 2A: Nearest Neighbor (BlastN vs. Genbank)			
SEQ ID	ACC'N	DESCRIP.	P VALUE
1038	AB015633.1	Homo sapiens mRNA for type II membrane protein, complete cds, clone:HP10481	7E-23
1039	U86751	Human nucleolar fibrillar center protein (ASE-1) mRNA, complete cds	0.019
1040	NM_001538.1	Homo sapiens heat shock transcription factor 4 (HSF4) mRNA >gi|1813425|dbj|D87673|D87673 Homo sapiens mRNA for heat shock transcription factor 4, complete cds	1E-57
1041	NM_001538.1	Homo sapiens heat shock transcription factor 4 (HSF4) mRNA >gi|1813425|dbj|D87673|D87673 Homo sapiens mRNA for heat shock transcription factor 4, complete cds	1E-57
1042	NM_001283.1	Homo sapiens clathrin-associated/assembly/adaptor protein, small 1 Homo sapiens mRNA for sigma1A subunit of AP-1 clathrin adaptor complex, complete cds	0
1043	L08265	Human skeletal muscle chloride channel (HUMCLC) gene, exon 7.	0.075
1044	U79731	Plasmodium berghei extrachromosomal plastid PB-1, ORF470 gene, partial cds, tRNA-Thr, large subunit ribosomal RNA, tRNA-Met, tRNA-Arg, tRNA-Val, tRNA-Arg, tRNA-Leu, tRNA Asn, tRNA-Ala, and small subunit ribosomal RNA genes...	0.037
1045	X64467	H.sapiens ALAD gene for porphobilinogen synthase	0.019
1046	M17374	X.laevis beta-globin mRNA, 5' UTR.	0.37
1047	AF144573.1	Mesocricetus auratus Mx-interacting protein kinase PKM mRNA, complete cds	2E-18
1048	U79260	Human clone 23745 mRNA, complete cds	7E-26
1049	Z23435	H. sapiens (D1S414) DNA segment containing (CA) repeat; clone AFM179xg5; single read	0.0007
1050	X95439	S.xylosus aroA, ccpA, acuC and acuA genes	0.014
1051	M32676	Human platelet glycoprotein IIIa, intron 10, fragment A.	0.011
1052	AB018284.1	Homo sapiens mRNA for KIAA0741 protein, complete cds	0.0009
1053	AF151843.1	Homo sapiens CGI-85 protein mRNA, complete cds	3E-33
1054	AF132966.1	Homo sapiens CGI-32 protein mRNA, complete cds	0
1055	NM_006466.1	Homo sapiens polymerase (RNA) III (DNA directed) (39kD) (RPC39) mRNA subunit (RPC39) mRNA, complete cds	4E-15
1056	AF086460	Homo sapiens full length insert cDNA clone ZD85A02	e-119
1057	AF036703	Caenorhabditis elegans cosmid T11F8	0.7
1058	AB018344.1	Homo sapiens mRNA for KIAA0801 protein, complete cds	0

Table 2A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1059	X81058	M.musculus tex261 mRNA	e-119
1060	AL035046.5	Human DNA sequence from clone 321I20 on chromosome 1q32.1-41 Contains GSSs, complete sequence	0.0001
1061	AF157814.1	Homo sapiens cAMP specific phosphodiesterase	0.00000002
1062	NM_002273.1	Homo sapiens keratin 8 (KRT8) mRNA keratin 8	e-120
1063	AF131739	Homo sapiens clone 25189 mRNA sequence, complete cds	0
1064	AL049702.1	Human gene from PAC 433G19, chromosome 1	0
1065	AL080125.1	Homo sapiens mRNA; cDNA DKFZp572P0920 (from clone DKFZp572P0920)	3E-19
1066	NM_003422.1	Homo sapiens zinc finger protein 42	2E-15
1067	Z22175	Caenorhabditis elegans cosmid K01F9, complete sequence [Caenorhabditis elegans]	2
1068	AF039690.1	Homo sapiens antigen NY-CO-8 (NY-CO-8) mRNA, partial cds	1E-37
1069	AL049702.1	Human gene from PAC 433G19, chromosome 1	0
1070	D63850	Mus musculus mRNA for hepatoma-derived growth factor, complete cds, strain:BALB/c	5E-50
1071	AB014603	Homo sapiens mRNA for KIAA0703 protein, complete cds	e-167
1072	NM_006371.1	Homo sapiens cartilage-associated protein sapiens mRNA for cartilage-associated protein (CASP)	0
1073	U91561	Rattus norvegicus pyridoxine 5'-phosphate oxidase mRNA, complete cds	e-136
1074	NM_003429.1	Homo sapiens zinc finger protein 85 (HPF4, HTF1) (ZNF85) mRNA >gi|1017721|gb|U35376|HSU35376 Human repressor transcriptional factor (ZNF85) mRNA, complete cds.	2E-51
1075	D63876	Human mRNA for KIAA0154 gene, partial cds	6E-54
1076	AB010059	Homo sapiens RBP56/hTAFII68 gene, exon 3, 4, 5	4E-80
1077	AB002584	Rattus norvegicus mRNA for beta-alanine-pyruvate aminotransferase, complete cds	0.00000002
1078	X85133	H.sapiens RBQ-1 mRNA	0
1079	AF074331.1	Homo sapiens PAPS synthetase-2 (PAPSS2) mRNA, complete cds	e-173

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
7	808943	(X82686) orf4 [bovine adenovirus type 2]	8.30E+00
8	3876268	(Z81067) similar to Zinc finger, C3HC4 type (RING finger) [Caenorhabditis elegans]	8.10E+00
9	2132973	probable membrane protein YPL058c - yeast	4.50E+00
11	2746799	(AF040643) No definition line found [Caenorhabditis elegans]	2.40E-01
12	1086865	(U41272) Similar to human leukocyte surface protein	1.80E-01
13	4680717	(AF132973) CGI-39 protein [Homo sapiens]	1.00E-07
14	5052588	(AF145649) BcDNA.GH08388	4.00E-09
15	2293303	(AF008220) YttA [Bacillus subtilis]	5.90E-02
16	3378132	(AF071502) brahma associated protein 155 kDa [Drosophila melanogaster]	4.20E-01
17	2224605	(AB002330) KIAA0332 [Homo sapiens]	7.30E-01
18	1439625	(U64598) weakly similar to S. cervisiae PTM1 precursor	1.30E+00
19	4758718	mitotic kinesin-like protein 1 PROTEIN-1 >gi|284312|pir||S28262 kinesin-related protein MKLP-1 - human >gi|34672|emb|CAA47628| (X67155) mitotic kinase-like protein-1 [Homo sapiens]	5.60E-01
20	399112	BETA-GALACTOSIDASE (LACTASE)	1.40E-01
26	3785995	(AC005499) unknown protein [Arabidopsis thaliana]	5.90E+00
27	5042442	(AC007789) putative CREB-binding protein [Oryza sativa]	3.50E+00
28	2501404	PUTATIVE ABC TRANSPORTER PERMEASE PROTEIN MJ0087 >gi|2127961|pir||G64310 hemin permease homolog - Methanococcus jannaschii >gi|1590869 (U67466) hemin permease (hemU) [Methanococcus jannaschii]	7.60E+00
29	1706771	5-EXO-ALCOHOL DEHYDROGENASE (FDEH) dehydrogenase [Pseudomonas putida]	6.00E+00
30	5102774	(AJ238893) acyl-CoA thioesterase [Mus musculus]	6.00E-11
31	121896	HISTONE H1.03 >gi|86287|pir||D28456 histone H1.03 - chicken >gi|211832 (M17021) 03 H1 protein [Gallus gallus]	4.80E-01
32	5689493	(AB029001) KIAA1078 protein [Homo sapiens]	1.00E-53
33	4063766	(D87895) chitinase [Emericella nidulans]	1.60E-02
34	4140029	(AB015438) alpha 1 type I collagen [Cynops pyrrhogaster]	2.70E-02
37	1182003	(X87904) putative [Homo sapiens]	2.70E+00
38	1072187	(U40941) coded for by C. elegans cDNA CEESB82F; coded for by C. elegans cDNA CEESE93F [Caenorhabditis elegans]	8.10E+00
39	2832671	(AL021712) hypothetical protein	1.40E+00
40	481043	bat2 protein - human >gi|29375|emb|CAA78744|	1.30E-01
41	1203952	(U49831) similar to D. melanogaster doublesex protein	4.80E+00

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
42	2708329	(AF038564) atrophin-1 interacting protein 4 [Homo sapiens]	8.20E+00
43	464522	TRANSCRIPTION INITIATION FACTOR IIF, ALPHA SUBUNIT (TFIIF-ALPHA) (TRANSCRIPTION INITIATION FACTOR RAP74) >gi|479869|pir||S35551 transcription factor IIF chain RAP74 - African clawed frog IIF subunit [Xenopus laevis]	3.30E-01
44	2384956	(AF022985) No definition line found [Caenorhabditis elegans]	2.00E-28
46	418745	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Crithidia oncopelti mitochondrion (SGC6) subunit 4 [Crithidia oncopelti]	8.40E+00
47	289825	(M81391) thrombin [Gallus gallus]	6.30E+00
48	1352968	HYPOTHETICAL 95.4 KD PROTEIN IN MAD2-RNR2 INTERGENIC REGION >gi|1077804|pir||S56801 hypothetical protein YJL029c - yeast (Saccharomyces cerevisiae) >gi|1008148|emb|CAA89320| (Z49304) ORF YJL029c [Saccharomyces cerevisiae]	2.80E+00
49	3878628	(Z93385) predicted using Genefinder; cDNA EST EMBL:D72583 comes from this gene; cDNA EST EMBL:D75500 comes from this gene [Caenorhabditis elegans]	6.00E-03
50	2384956	(AF022985) No definition line found [Caenorhabditis elegans]	3.00E-25
54	5262605	(AL080150) hypothetical protein [Homo sapiens]	2.10E+00
59	4680695	(AF132962) CGI-28 protein [Homo sapiens]	2.40E-01
60	961446	(D63877) KIAA0157 gene product is novel.	1.20E+00
61	3882321	(AB018343) KIAA0800 protein [Homo sapiens]	1.00E-69
62	2864624	(AL021811) putative protein [Arabidopsis thaliana]	1.40E-01
65	1655907	(U65891) protein tyrosine phosphatase CRYP-2 [Gallus gallus]	2.00E+00
66	3983370	(AF102521) olfactory receptor B12 [Mus musculus]	1.80E-01
67	103624	collagen alpha 2 chain - sea urchin 2-alpha collagen precursor (COLL 2-alpha) [Paracentrotus lividus]	1.50E+00
68	3881789	(Z68302) predicted using Genefinder; similar to Pumilio-family RNA binding domains (aka PUM-HD, Pumilio homology domain) (3 domains); cDNA EST EMBL:M89238 comes from this gene; cDNA EST EMBL:D73612 comes from this gene; cDNA ES...	2.50E-01
70	1326281	(U58732) F48D6.2 gene product [Caenorhabditis elegans]	3.40E+00
73	3540219	(D87686) KIAA0017 protein [Homo sapiens]	8.00E-70

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
74	2315742	(AF016681) contains similarity to a sperm coat polysaccaride domain [Caenorhabditis elegans]	1.60E+00
78	5453948	protein phosphatase 1, regulatory subunit 6 >gi|3805797|emb|CAA77081| (Y18206) serine-threonine specific protein phosphatase [Homo sapiens]	6.50E+00
80	1161051	(L39922) efflux protein [Mycobacterium tuberculosis]	8.20E+00
81	4505279	5-methyltetrahydrofolate-homocysteine methyltransferase reductase >gi|2981303 (AF025794) methionine synthase reductase [Homo sapiens]	2.80E+00
82	1722711	MAJOR CAPSID PROTEIN L1 >gi|1020201 type 24]	1.20E+00
84	5453379	(AF155124) bacterial-induced peroxidase precursor [Gossypium hirsutum]	6.30E+00
85	3329139	(AE001339) ABC Transporter Membrane Protein [Chlamydia trachomatis]	1.20E+00
86	222416	(D10453) coat protein [Pea seed-borne mosaic virus]	4.50E+00
87	1778160	(U67304) 70 kDa S6 kinase [Drosophila melanogaster]	2.60E+00
88	1943947	(U90126) ABC transporter [Bos taurus]	2.60E+00
89	3851586	(AF092564) chromosome-associated protein-C [Homo sapiens]	2.00E-03
91	119711	EXTENSIN PRECURSOR carota] >gi|224686|prf||1111211A extensin [Daucus carota]	2.00E-03
93	728838	!!!! ALU SUBFAMILY SX WARNING ENTRY	6.10E-01
95	4406632	(AF131801) Unknown [Homo sapiens]	2.00E-04
97	4585699	(AJ228139) LEKTI precursor [Homo sapiens]	9.00E-52
99	3249026	(AF070067) unknown [Escherichia coli]	7.70E-01
100	4210358	(AL031073) dJ142F18.1 (similar to melanoma-associated antigen) [Homo sapiens]	1.70E-02
101	2137074	ribosomal transcription factor UBF2 - Chinese hamster	7.00E-06
103	3327062	(AB014524) KIAA0624 protein [Homo sapiens]	1.00E-37
106	3328339	(AF075241) prepro-orexin [Sus scrofa]	4.80E+00
108	1055163	(U40029) Contains similarity to Pfam domain: PF01060 (Worm_family_2), Score=203.8, E-value=8.6e-58, N=1 [Caenorhabditis elegans]	7.90E+00

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
109	113333	METHYLPHOSPHOTRIESTER-DNA ALKYLTRANSFERASE >gi|279475|pir||XUBSMM methylphosphotriester-DNA methyltransferase (EC 2.1.1.-) adaA [Bacillus subtilis] >gi|2632448|emb|CAB11957| transcriptional regulator (AraC/XylS family) [Bacillus subtilis] >gi|2632466|emb|CAB11974| (Z99105) methylphosphotriester-DNA alkyltransferase and transcriptional regulator (AraC/XylS family) [Bacillus subtilis] >gi|3599599|dbj|BAA33074| (AB006424) METHYLPHOSPHOTRIESTER-DNA ALKYLTRANSFERASE [Bacillus subtilis]	7.80E+00
110	2143767	glycoprotein - rat >gi|986943 (L08134) glycoprotein [Rattus norvegicus] norvegicus]	2.00E-02
112	2905979	(AF015678) virulence determinant [African swine fever virus]	2.00E+00
113	4867999	(AF078164) Ku70-binding protein	4.00E-60
114	3820909	(AJ010642) Dof protein [Drosophila melanogaster]	1.90E+00
115	2291155	(AF016418) No definition line found [Caenorhabditis elegans]	8.10E+00
120	632500	(U17394) polyadenylation factor 64 kDa subunit [Xenopus laevis]	3.60E+00
121	87792	Ig gamma-3 chain C region (allotype G3m(b)) - human >gi|577056|emb|CAA27268| (X03604) C gamma 3 [Homo sapiens]	1.60E+00
122	3043572	(AB011096) KIAA0524 protein [Homo sapiens]	5.00E-04
125	1729859	TUBULIN GAMMA CHAIN gamma tubulin-like protein [Saccharomyces cerevisiae]	3.50E+00
126	2340169	(AF015783) telomerase reverse transcriptase 1	2.70E+00
127	2131446	hypothetical protein YDR362c - yeast	7.90E-02
130	1616770	(U70731) putative poly(A)-binding protein FabM	1.10E+00
131	3287688	(AC003979) Contains similarity to ycf37 gene product gb|1001425 from Synechocystis sp. genome gb|D63999. ESTs gb|T43026, gb|R64902, gb|Z18169 and gb|N37374 come from this gene. [Arabidopsis thaliana]	8.00E-03
134	1131444	(U42580) PBCV-1 glucosamine synthetase	1.70E+00
135	134952	STREPTOTHRICIN ACETYLTRANSFERASE streptothricin acetyl-transferase (AA 1-174) streptothricin-acetyl-transferase (AA 1-174) acetyltransferase [Transposon Tn7] >gi|2708491 (U84739) streptothricin resistance protein [synthetic construct] acetyltransferase 3' [Cloning vector pSB11]	5.40E-01

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
136	3452684	(D87054) 2-heptaprenyl-1,4-naphthoquinone methyltransferase [Bacillus stearothermophilus]	4.00E-03
137	5360129	(AF155117) NY-REN-62 antigen	8.00E-53
143	1363109	collagen alpha 1(XVIII) chain precursor long form - mouse (fragment) >gi|618430 (U11637) alpha-1 type XVIII collagen precursor [Mus musculus]	3.70E+00
145	113671	!!!! ALU CLASS F WARNING ENTRY !!!!	1.40E+00
146	3874135	(Z54342) similar to acid phosphatase elegans]	7.00E-22
148	1279390	(X97329) HER-1 protein [Danio rerio]	7.50E+00
149	4557639	orexin receptor 2 >gi|2897128 receptor [Homo sapiens]	6.20E+00
150	4262630	(AF125963) No definition line found	3.20E+00
152	102129	H+-transporting ATP synthase (EC 3.6.1.34) protein 6 - Trypanosoma brucei mitochondrion (SGC6)	2.80E+00
156	996018	(X91637) BRG1 protein [Gallus gallus]	5.70E+00
157	3158498	(AF067622) Contains similarity to Pfam domain: PF00628 (PHD), Score=36.7, E-value=1.7e-07, N=2	2.70E-02
158	1117913	(U40223) uridine nucleotide receptor [Homo sapiens]	2.70E+00
162	5430752	(AC007504) Hypothetical Protein	3.80E-02
167	226120	vicilin gene B [Saguinus oedipus]	8.30E+00
169	4557335	aspartoacylase (aminoacylase 2) aspartoacylase - human >gi|455834|bbs|140585 (S67156) aspartoacylase, ASP [human, kidnay, Peptide, 313 aa]	7.30E+00
170	5031129	(AF082859) lungkine [Mus musculus]	8.60E+00
173	4678899	(AL049707) putative large glycine/alanine rich protein [Streptomyces coelicolor]	2.10E-01
174	728831	!!!! ALU SUBFAMILY J WARNING ENTRY	2.00E-02
175	3004981	(AF039652) ribonuclease H type II [Homo sapiens]	2.00E-27
177	2911366	(AF041047) NADPH HC toxin reductase [Zea mays]	9.60E-02
178	2217964	(Z50798) p52 [Gallus gallus]	1.00E-12
179	3873707	(Z73102) Similarity to B.subtilis DNAJ protein (SW:DNAJ_BACSU); cDNA EST yk437a1.5 comes from this gene [Caenorhabditis elegans]	3.00E-19
180	3043658	(AB011139) KIAA0567 protein [Homo sapiens]	2.00E-03
182	5689451	(AB028980) KIAA1057 protein [Homo sapiens]	7.00E-10
183	728831	!!!! ALU SUBFAMILY J WARNING ENTRY	5.00E-03
186	2588623	(AC003083) mitochondrial carrier protein-like; similar to Q09461 (PID:g2497990) [Homo sapiens]	3.00E-69
187	1669601	(D88747) AR401 [Arabidopsis thaliana]	2.00E-20
190	126296	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG protein [Nycticebus coucang]	3.20E-01

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
191	3294180	(Z99129) dJ425C14.2 (Placental protein DIFF33 LIKE) [Homo sapiens]	4.00E-20
192	5030439	(AC007766) R26610_1 [Homo sapiens]	7.00E-56
193	4507375	tubulin-specific chaperone e	7.00E-05
195	1698455	(U49974) mariner transposase [Homo sapiens]	2.00E-05
196	1709285	PUTATIVE PYRUVATE-FLAVODOXIN OXIDOREDUCTASE >gi|1006618|dbj|BAA10774| (D64005) pyruvate oxidoreductase [Synechocystis sp.]	8.00E+00
197	3878584	(Z77667) cDNA EST EMBL:C08125 comes from this gene; cDNA EST EMBL:C09753 comes from this gene	2.00E-04
198	1658503	(U75467) Atu [Drosophila melanogaster]	2.00E-44
199	2655422	(AF035530) CDC37 [Gallus gallus]	2.00E-09
200	4557817	3-oxoacid CoA transferase precursor; succinyl-CoA:3-ketoacid-CoA transferase precursor >gi|2492998|sp|P55809|SCOT_HUMAN SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR transferase precursor [Homo sapiens]	3.20E+00
201	4506403	UNKNOWN >gi|4063890 (AF095448) putative G protein-coupled receptor [Homo sapiens]	4.00E-35
202	3880146	(Z68319) Similarity to Human hnRNP F protein (PIR Acc. No. S43484); cDNA EST EMBL:D34218 comes from this gene; cDNA EST EMBL:D37248 comes from this gene; cDNA EST EMBL:D71817 comes from this gene; cDNA EST EMBL:D74531 comes fro... hnRNP F protein (PIR Acc. No. S43484); cDNA EST EMBL:D34218 comes from this gene; cDNA EST EMBL:D37248 comes from this gene; cDNA EST EMBL:D71817 comes from this gene; cDNA EST EMBL:D74531 comes fro...	1.00E-01
203	3877198	(Z69903) predicted using Genefinder; Similarity to Rat casein kinase I (SW:KC1D_RAT); cDNA EST EMBL:D65322 comes from this gene; cDNA EST EMBL:D68704 comes from this gene; cDNA EST yk475f2.5 comes from this gene [Caenorhabditis...	1.20E+00
204	987050	(X65335) lacZ [Cloning vector pSV-beta-Galactosidase Control]	4.00E-06
205	1438677	(U62376) envelope protein [Simian immunodeficiency virus]	3.60E+00
206	2193870	(D84391) reverse transcriptase [Mus musculus]	6.00E-06
207	5032069	spliceosome-associated protein ASSOCIATED PROTEIN 49 (SAP 49) (SF3B53) SAP-49 - human >gi|556217 (L35013) spliceosomal protein	2.20E+00

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
209	5032069	spliceosome-associated protein ASSOCIATED PROTEIN 49 (SAP 49) (SF3B53) SAP-49 - human >gi|556217 (L35013) spliceosomal protein	2.10E+00
210	729264	CYTOCHROME B >gi|625356|pir||S43269 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - humpback whale mitochondrion (SGC1)	4.60E+00
211	2088713	(AF003139) Similar to cuticular collagen [Caenorhabditis elegans]	4.20E-01
212	2276366	(Z97992) putative glucan synthase	8.40E+00
213	4589684	(AB023234) KIAA1017 protein [Homo sapiens]	2.00E-65
214	4504739	ITBA1 protein	1.00E-10
217	1703364	TRANSCRIPTIONAL REGULATORY PROTEIN ARAB >gi|995682|emb|CAA62739| (X91393) abaB	7.50E-01
218	2769595	(Y16135) 5HT2B receptor [Canis familiaris]	8.20E+00
219	3002527	(AF010144) neuronal thread protein AD7c-NTP [Homo sapiens]	5.00E-05
221	1463014	(U08794) envelope glycoprotein [Human immunodeficiency virus type 1]	7.70E+00
222	1709230	NBL4 PROTEIN >gi|543191|pir||JU0188 band 4.1 superfamily member protein - mouse	3.00E-23
223	120223	FK506-BINDING PROTEIN (FKBP-12) FK506-binding protein - mouse >gi|50971|emb|CAA42762| musculus]	1.00E-19
224	987050	(X65335) lacZ [Cloning vector pSV-beta-Galactosidase Control]	2.00E-15
225	2558516	(AJ001119) Rab5 GDP/GTP exchange factor, Rabex5 [Bos taurus]	2.00E-36
226	2558501	(D63850) hepatoma-derived growth factor	7.00E-30
227	5410326	(AF106680) RNA helicase [Homo sapiens]	2.00E-45
228	631772	TEG-261 protein - mouse	5.00E-48
229	3851492	(AF041853) kinesin family member protein KIF3A [Homo sapiens]	3.00E-56
231	4504983	lectin, galactoside-binding, soluble, 3 (galectin 3) (NOTE: redefinition of symbol) BINDING PROTEIN 35) (CBP 35) (LAMININ-BINDING PROTEIN) galactoside-binding - human >gi|179531 (M57710) IgE-binding protein [Homo sapiens] >gi|186922 (M36682) laminin-binding protein [Homo sapiens]	6.00E+00
240	3873717	(Z81453) predicted using Genefinder	3.70E+00
241	3413884	(AB007930) KIAA0461 perotein [Homo sapiens]	3E-48
242	1834503	(Z72496) mucin MUC5B [Homo sapiens]	4.50E-01

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
243	131442	PRESTALK PROTEIN PRECURSOR mold (Dictyostelium discoideum)	1.2
244	3834629	(AF094519) diaphanous-related formin; p134 mDia2 [Mus musculus]	5E-54
245	2996650	(AC004493) KIAA0324 [Homo sapiens]	0.05
246	731604	HYPOTHETICAL 130.0 KD PROTEIN IN SNF6-SPO11 INTERGENIC REGION >gi|626572|pir||S46837 hypothetical protein YHL023c - yeast (Saccharomyces cerevisiae) >gi|2289893 (U11582) No definition line found [Saccharomyces cerevisiae]	0.036
247	5670007	(AF156102) ELL complex EAP30 subunit	5.00E-66
248	1122431	(X92968) protein SIC [Streptococcus pyogenes]	0.001
249	462022	ER LUMEN PROTEIN RETAINING RECEPTOR falciparum >gi|398385|emb|CAA81128| (Z26043) ERD2	8.80E+00
250	3850153	(AL033396) cytochrome P450 [Candida albicans]	8.8
251	2315228	(Z98260) hypothetical protein Rv1227c	1.00E+00
252	3800952	(AF100657) Contains similarity to Pfam domain: PF00614 (PLDc), Score=13.8, E-value=0.2, N=1	5.00E-24
253	4507145	UNKNOWN >gi|3873216 (AF065485) sorting nexin 4 [Homo sapiens]	3.00E-51
258	5410355	(AF125392) insulin induced protein 2 [Homo sapiens]	2.10E+00
261	2905647	(AF045245) D-arabinitol kinase [Klebsiella pneumoniae]	6.5
262	4928673	(AF136343) Cul-1 [Drosophila melanogaster]	6.50E+00
263	3876644	(Z81526) predicted using Genefinder; cDNA EST EMBL:D36935 comes from this gene; cDNA EST EMBL:D33960 comes from this gene; cDNA EST EMBL:C12255 comes from this gene; cDNA EST EMBL:C10859 comes from this gene; cDNA EST EMBL:C1...	6.20E+00
264	1808621	(X94355) D18L [Cowpox virus]	3.70E+00
265	628784	plasmid copy number control protein - Escherichia coli >gi|473802|dbj|BAA05591| (D26562) coli]	2.90E+00
266	4505727	peroxisomal biogenesis factor 3 PROTEIN PEX3 (PEROXIN-3) >gi|3336882|emb|CAA04879| sapiens] >gi|4218426|emb|CAA08904| (AJ009866) Pex3p	4.00E-59
268	4322053	(AF071242) homeobox protein [Danio rerio]	3.50E+00
273	5105878	(AP000063) 194aa long hypothetical protein [Aeropyrum pernix]	6.50E+00

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
274	3877495	(Z48583) cDNA EST EMBL:T00483 comes from this gene; cDNA EST EMBL:D64526 comes from this gene; cDNA EST EMBL:D65147 comes from this gene; cDNA EST EMBL:D68484 comes from this gene; cDNA EST EMBL:D67548 comes from this gene; c... >gi|3879229|emb|CAA88749| EST EMBL:D64526 comes from this gene; cDNA EST EMBL:D65147 comes from this gene; cDNA EST EMBL:D68484 comes from this gene; cDNA EST EMBL:D67548 comes from this gene; cDN...	6.7
275	3322592	(AE001211) T. pallidum predicted coding region TP0311	1.90E+00
276	1176495	HYPOTHETICAL 79.4 KD PROTEIN IN PRP16-SRP40 INTERGENIC REGION >gi|486577|emb|CAA82169|	2.1
277	4691726	(AF124490) ARF GTPase-activating protein GIT1 [Homo sapiens]	3E-68
278	3702844	(AF069051) pituitary tumor transforming gene protein	5.80E+00
279	3882163	(AB018264) KIAA0721 protein [Homo sapiens]	6.00E-59
280	1072187	(U40941) coded for by C. elegans cDNA CEESB82F; coded for by C. elegans cDNA CEESE93F [Caenorhabditis elegans]	7.4
282	3322397	(AE001198) T. pallidum predicted coding region TP0130	1.80E+00
283	3417412	(AL031261) putative superoxide dismutase	2.9
284	2245121	(Z97343) hypothetical protein	0.45
285	2924311	(AJ000882) steroid receptor coactivator 1e	8.6
286	3323285	(AE001264) ABC transporter, ATP-binding protein	8.5
287	4981435	(AE001755) hypothetical protein	2.9
288	4826818	LIM and senescent cell antigen-like domains 1 >gi|1346721|sp|P48059|PINC_HUMAN PINCH PROTEIN (PARTICULARY INTERESTING NEW CYS-HIS PROTEIN) >gi|631281|pir||JC2324 LIM protein - human	6.5
291	104506	troponin T, fast skeletal muscle, embryonic alpha (clone 501) - Japanese quail >gi|213628 (M26599) troponin T [Coturnix coturnix]	4.8
292	2736462	(AF039048) similar to cdc25-like M-phase inducer phosphatases [Caenorhabditis elegans]	2.8
293	1708966	SPERM MITOCHONDRIAL CAPSULE SELENOPROTEIN (MCS)	0.74
294	4914378	(AC007584) hypothetical protein [Arabidopsis thaliana]	3E-10
295	5524931	(AL096842) hypothetical protein [Homo sapiens]	3E-64

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
296	128155	LOW-AFFINITY NERVE GROWTH FACTOR RECEPTOR PRECURSOR (NGF RECEPTOR) (GP80-LNGFR) (P75 ICD) affinity - chicken	0.00001
298	4205113	(AF000520) cell wall invertase [Fragaria x ananassa]	2.7
299	3702106	(AL031765) MSK (EC 2.7.1.-) (HRT-20) (MYOCARDIAL SNF1-LIKE KINAS...	0.79
300	4325123	(AF119361) unknown [Frankia sp. EuIK1]	2.8
301	417737	MITOCHONDRIAL RIBOSOMAL PROTEIN S14 polymorpha=liverwort, Peptide Mitochondrial, 99 aa]	3.6
302	2289030	(U53564) N-terminal region of the protein [Mus musculus]	2.8
303	2318003	(U97553) unknown [murine herpesvirus 68]	0.037
304	3123176	HYPOTHETICAL 43.1 KD TRP-ASP REPEATS CONTAINING PROTEIN K04G11.4 IN CHROMOSOME X Genefinder; Similarity to C.elegans Guanine nucleotide binding protein (WP:C14B1.4) [Caenorhabditis elegans]	6E-09
305	5106956	(AF113615) FH1/FH2 domain-containing protein FHOS [Homo sapiens]	1E-51
311	5059323	(AF151522) hairy and enhancer of split related-1 [Homo sapiens]	0.31
312	728831	!!!! ALU SUBFAMILY J WARNING ENTRY	4.7
313	4226073	(AF125443) contains similarity to S. pombe phosphatidyl synthase (GB:Z28295) [Caenorhabditis elegans]	2E-23
317	2822320	(AF016485) ORF H0532 [Halobacterium sp. NRC-1] >gi|2822445|gb|AAC82951.1| (AF016485) ORF H1831	7.9
318	2983553	(AE000721) major facilitator family transporter [Aquifex aeolicus]	3.5
319	4689225	(AF118379) gamma-tubulin ring protein Dgrip84 [Drosophila melanogaster]	0.23
322	1326337	(U58746) coded for by C. elegans cDNA yk3b11.5; coded for by C. elegans cDNA yk13g1.5; coded for by C. elegans cDNA yk3b11.3; coded for by C. elegans cDNA CEESR37F; coded for by C. elegans cDNA yk13g1.3; Similar to phospholipase. [Caenorhab...	4.5
323	2708738	(AC003952) hypothetical protein [Arabidopsis thaliana]	6E-10
326	4929629	(AF151838) CGI-80 protein [Homo sapiens]	8.1
327	4050073	(AF103731) putative glycolipid transfer protein [Homo sapiens]	2E-38
328	1905892	(L39835) Na/Ca exchange protein [Drosophila melanogaster]	0.14
329	4972120	(AL078579) putative protein [Arabidopsis thaliana]	2E-08

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
330	4884202	(AL049953) hypothetical protein [Homo sapiens]	1E-39
332	2981631	(AB012223) ORF2 [Canis familiaris]	0.098
333	1723611	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN GLVC-LIPB INTERGENIC REGION subtilis] >gi|2633149|emb|CAB12654| (Z99108) similar to transcriptional regulator (AraC/XylS family) [Bacillus subtilis]	9.9
334	3881873	(Z83246) predicted using Genefinder; cDNA EST EMBL:M79771 comes from this gene [Caenorhabditis elegans]	1.3
335	3641352	(AF091234) putative transcription factor [Mus musculus]	2E-43
336	3874427	(Z78416) predicted using Genefinder; Similarity to S.pombe RAD18 gene (TR:E198069); cDNA EST CEESX52R comes from this gene; cDNA EST EMBL:D32785 comes from this gene; cDNA EST EMBL:D35528 comes from this gene; cDNA EST EMBL:D37...	0.0000006
341	1492037	(U60315) MC094R [Molluscum contagiosum virus subtype 1]	8.5
342	3201625	(AC004669) hypothetical protein [Arabidopsis thaliana]	6.2
343	5103944	(AP000059) 216aa long hypothetical protein [Aeropyrum pernix]	3.8
344	139140	RNA REPLICATION PROTEIN (165 KD PROTEIN) (ORF 1) [CONTAINS: RNA-DIRECTED RNA POLYMERASE RNA-replicating protein [Potato virus X] >gi|309911	5.9
345	1085126	juvenile hormone esterase-related protein - cabbage looper	4.9
346	1613846	(U71440) polyprotein [Rice tungro spherical virus]	0.73
354	60900	(X03614) alternative form of op-6 (aa 1-1980) [Human parainfluenza virus 1]	0.35
355	3287370	(AC002397) B [Mus musculus]	0.003
356	2622601	(AE000909) serine/threonine protein kinase related protein [Methanobacterium thermoautotrophicum]	2E-10
358	130489	STRUCTURAL POLYPROTEIN [CONTAINS: MAJOR STRUCTURAL PROTEIN VP2; NONSTRUCTURAL PROTEIN VP4; MINOR STRUCTURAL PROTEIN VP3] >gi|75451|pir||GNXSOH genome polyprotein - infectious bursal disease virus structural polyprotein [Infectious bursal disease virus]	9.7
359	2996337	(AF053947) CobT homolog [Yersinia pestis]	0.86
360	3644048	(AF091395) Trio isoform [Homo sapiens]	7.1
363	3845280	(AE001418) hypothetical protein [Plasmodium falciparum]	0.8

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
364	1208844	(U49956) coded for by C. elegans cDNA yk57d5.5; coded for by C. elegans cDNA cm20e2; coded for by C. elegans cDNA cm06f2 [Caenorhabditis elegans]	4
366	3193245	(AF068709) No definition line found [Caenorhabditis elegans]	2.9
368	5262568	(AL080129) hypothetical protein [Homo sapiens]	1E-35
371	3261730	(Z92774) nhoA [Mycobacterium tuberculosis]	3.5
373	2131482	hypothetical protein YDR426c - yeast	7.7
374	2146218	hypothetical protein E30_orf352 - Mycoplasma pneumoniae (SGC3) (ATCC 29342) >gi|1673872 (AE000021) Mycoplasma pneumoniae, E30_orf352 Protein [Mycoplasma pneumoniae]	7.8
375	2911866	(AF047660) contains similarity to steroid/thyroid/retinoic nuclear hormone receptors; contains similarity to C4-type zinc fingers	2.7
377	3878117	(Z49068) mitochondrial carrier protein	5.7
378	2804500	(AF043706) contains similarity to granulins [Caenorhabditis elegans]	0.18
380	4154882	(AE001471) ATP-DEPENDENT ZINC METALLOPEPTIDASE	4.7
381	2047346	(AF000198) Similar to cuticular collagen [Caenorhabditis elegans]	0.31
382	135454	TUBULIN BETA-2 CHAIN Emericella nidulans >gi|168107 (M17520) beta-tubulin beta [Emericella nidulans]	1.5
387	3256691	(AP000001) 128aa long hypothetical protein [Pyrococcus horikoshii]	3.6
388	4033606	(AB008227) Extensin [Adiantum capillus-veneris]	0.33
391	3036835	(AJ003243) bradykinin B2 receptor [Cavia porcellus]	7.9
392	5306171	(AF160864) NADH dehydrogenase subunit 4 [Tetrahymena pyriformis]	1.6
393	2131472	hypothetical protein YDR409w - yeast CAI: 0.12 [Saccharomyces cerevisiae]	0.43
394	418745	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Crithidia oncopelti mitochondrion (SGC6) subunit 4 [Crithidia oncopelti]	3.3
398	422408	cyclodiene insecticide resistance protein - yellow fever mosquito >gi|881590 (U28803) GABA receptor subunit [Aedes aegypti]	1.1
400	4050089	(AF109907) hypothetical protein [Homo sapiens]	1.6

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
401	3875400	(Z73906) cDNA EST EMBL:M88866 comes from this gene [Caenorhabditis elegans]	2.1
402	3874821	(Z35641) cDNA EST yk273d8.5 comes from this gene [Caenorhabditis elegans]	9E-10
404	4493761	(AL034368) predicted using hexExon; L779.2, Hypothetical protein, len: 4125 aa [Leishmania major]	6.6
407	2598627	(AJ000870) histidine kinase [Streptococcus gordonii]	6
408	2842531	(AB004291) gamma-subunit of enolase	3.5
409	728836	!!!! ALU SUBFAMILY SP WARNING ENTRY	0.37
415	728832	!!!! ALU SUBFAMILY SB WARNING ENTRY	0.95
416	2088675	(AF003131) C. elegans UNC-89 (GB:U33058) (NID:g1160355)	1.2
417	102189	myosin I, high molecular weight - Acanthamoeba sp	0.0005
418	2072961	(U93568) putative p150 [Homo sapiens]	0.008
422	1705706	P-TYPE CALCIUM CHANNEL ALPHA-1 SUBUNIT (RBA-I) >gi|111447|pir||A41098 calcium channel protein alpha-1 chain isoform A - rat >gi|203111 norvegicus]	3.6
424	1781316	(Y10290) formamidopyrimidine-DNA glycosylase [Synechococcus elongatus]	4.9
425	1183033	(D63821) polyprotein [Hepatitis C virus]	7.5
426	3347920	(AF075261) orphan transporter [Mus musculus]	2.9
428	1333929	(X66285) HC1 ORF [Mus musculus]	0.086
429	3121994	DNAJ PROTEIN japonicum]	1.2
431	5689523	(AB029016) KIAA1093 protein [Homo sapiens]	0.001
432	4887240	(AF064564) WSB1 protein [Fugu rubripes] rubripes]	0.013
433	130553	RNA REPLICASE POLYPROTEIN 2.7.7.48) - Ononis yellow mosaic virus >gi|332574 virus]	0.3
435	897917	(U28249) 11kD protein [Homo sapiens]	0.25
441	2072958	(U93567) putative p150 [Homo sapiens]	0.002
444	728831	!!!! ALU SUBFAMILY J WARNING ENTRY	0.008
446	4210496	(U61384) GAS41 protein [Homo sapiens]	9E-59
447	1002672	(U30261) G protein beta subunit-like; Method: conceptual translation supplied by author [Schistosoma mansoni]	1E-31
448	4680703	(AF132966) CGI-32 protein [Homo sapiens]	6E-67
450	1255371	(U53147) coded for by C. elegans cDNA yk34a9.5; coded for by C. elegans cDNA yk34a9.3; Similar to guanylate kinase. [Caenorhabditis elegans]	4E-23
451	1078718	reverse transcriptase - Trypanosoma cruzi transcriptase [Trypanosoma cruzi]	0.91
452	728832	!!!! ALU SUBFAMILY SB WARNING ENTRY	0.082
453	106323	hypothetical protein (L1H 5' region) - human	0.58

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
454	1345693	CHLORAMPHENICOL ACETYLTRANSFERASE acetyltransferase, CAT [Vibrio anguillarum=pJA7324, Peptide Plasmid, 216 aa] [Vibrio anguillarum]	8.4
455	3413884	(AB007930) KIAA0461 perotein [Homo sapiens]	8E-78
456	2072972	(U93572) putative p150 [Homo sapiens]	0.003
458	1504042	(D86984) similar to yeast adenylate cyclase (S56776) [Homo sapiens]	7E-10
460	729774	HEAT SHOCK FACTOR PROTEIN HSF30 STRESS TRANSCRIPTION FACTOR) >gi|100265|pir||S25480 heat shock transcription factor HSF30 - Peruvian tomato transcription factor HSF30 [Lycopersicon peruvianum]	8.1
462	728831	!!!! ALU SUBFAMILY J WARNING ENTRY	0.006
464	1351218	TESTIN 2 (TES2) [CONTAINS: TESTIN 1 (TES1)] >gi|2137810|pir||I48842 testin - mouse	7.8
465	4160198	(AL008583) dJ327J16.3 (novel CHROMObox family protein) [Homo sapiens]	1E-20
466	4454690	(AF070657) glutathione S-transferase subunit 13 homolog [Homo sapiens]	2E-25
467	3873871	(Z81030) similar to citrate lyase beta chain; cDNA EST yk302b4.5 comes from this gene	3E-41
468	2690136	(AE000788) conserved hypothetical protein [Borrelia burgdorferi]	4.7
469	3327192	(AB014589) KIAA0689 protein [Homo sapiens]	0.000006
470	121654	GASTRULA-SPECIFIC PROTEIN 17 African clawed frog >gi|64733|emb|CAA28842| (X05215) GS17 gene product (AA 1 - 147) [Xenopus laevis]	0.9
471	2506774	KERATIN, TYPE II CYTOSKELETAL 8	2E-42
472	4768838	(AF116910) putative ribonuclease III [Homo sapiens]	6E-74
473	4406551	(AF131739) Unknown [Homo sapiens]	2E-54
474	4454704	(AF070664) HSPC008 [Homo sapiens] protein [Homo sapiens]	3E-39
475	4678836	(AL049701) hypothetical protein [Homo sapiens]	6E-43
477	3025319	ZINC FINGER PROTEIN 195 >gi|2384653 sapiens]	3E-11
479	3327098	(AB014542) KIAA0642 protein [Homo sapiens]	5E-20
480	2506062	(D85196) cut4+ [Schizosaccharomyces pombe]	4.7
481	220579	(D00570) open reading frame (196 AA) [Mus musculus]	1.7
482	4758756	nucleosome assembly protein 1-like 1 >gi|1709337|sp|P55209|NPL1_HUMAN NUCLEOSOME ASSEMBLY PROTEIN 1-LIKE 1 (NAP-1 RELATED PROTEIN)	2E-26
484	1778432	(U79660) Treacher Collins syndrome [Homo sapiens]	2.9

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
485	4507455	transferrin receptor 2 sapiens]	3E-35
488	1184790	(U46068) von Ebner minor salivary gland protein [Mus musculus]	0.065
489	3876562	(Z81074) Similarity to Soybean 3-methylcrotonyl-CoA carboxylase (TR:Q42777); cDNA EST EMBL:M75819 comes from this gene; cDNA EST EMBL:M89099 comes from this gene; cDNA EST EMBL:D32737 comes from this gene; cDNA EST EMBL:D32763 ...	7E-41
490	746552	(U23523) F53A9.1 gene product [Caenorhabditis elegans]	6.7
491	2981631	(AB012223) ORF2 [Canis familiaris]	0.037
492	1401210	(U58510) putative RNA polymerase II subunit	4.8
493	3170180	(AF039690) antigen NY-CO-8 [Homo sapiens]	0.26
496	1778838	(U83113) INS-1 winged-helix homolog [Homo sapiens]	2.8
497	549779	PUTATIVE MYCOCEROSYL TRANSFERASE IN MAS 5'REGION >gi|322248|pir||A44110 orf I 5' of mas - Mycobacterium tuberculosis >gi|149979 (M95808) ORF	8.2
498	3877493	(Z48583) similar to ATPases associated with various cellular activities (AAA); cDNA EST EMBL:Z14623 comes from this gene; cDNA EST EMBL:D75090 comes from this gene; cDNA EST EMBL:D72255 comes from this gene; cDNA EST yk200e4....	3E-14
500	3879937	(Z68220) T20D3.3 [Caenorhabditis elegans]	0.0000003
501	1170551	MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 2 >gi|1078046|pir||S53952 proteinase 2 precursor, mitochondrial inner membrane - yeast	4E-13
502	4210989	(AF121781) unknown [Homo sapiens]	0.007
503	4826454	(Z93241) dJ222E13.3.2 (PUTATIVE partial isoform 2) [Homo sapiens]	2E-46
504	5381426	(AF159046) SPANK-1 [Rattus norvegicus]	0.12
505	3687833	(AF069737) notchless [Xenopus laevis]	1E-65
506	2558501	(D63850) hepatoma-derived growth factor	3E-24
507	1061310	(M98326) valyl-tRNA synthetase [Homo sapiens]	2E-17
508	4503179	gene encoding a protein with coiled-coil alpha-helical domains protein [Homo sapiens]	3E-35
509	4096591	(U33460) DNA-directed RNA polymerase I, largest subunit [Homo sapiens]	6.3
510	4836515	(AF124788) WS-3 protein [Mus musculus]	5E-10
511	4507867	vessicle-associated membrane protein (VAMP)-associated protein of 33 kDa >gi|3320446 sapiens]	9E-33
512	5262560	(AL080125) hypothetical protein [Homo sapiens]	1E-41

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
513	134039	SMALL NUCLEAR RIBONUCLEOPROTEIN SM D1 (SNRNP CORE PROTEIN D1) (SM-D1) (SM-D AUTOANTIGEN) Sm-D [Homo sapiens] >gi|1256741 (M58558) Sm-D autoantigen [Mus musculus]	6E-13
514	4165247	(AL021397) dJ69E11.3 (Yeast YPR037W and worm C02C2.6 predicted proteins LIKE) [Homo sapiens] protein [Homo sapiens]	2E-52
515	3327220	(AB014603) KIAA0703 protein [Homo sapiens]	5E-53
516	4506599	ribosomal protein L13 L13 (BREAST BASIC CONSERVED PROTEIN 1) sapiens]	0.0000003
517	3877201	(Z70780) cDNA EST yk465d10.3 comes from this gene; cDNA EST yk465d10.5 comes from this gene; cDNA EST yk481d9.5 comes from this gene [Caenorhabditis elegans]	0.00002
519	5453601	cartilage-associated protein cartilage-associated protein (CASP) [Homo sapiens]	8E-70
520	4633085	(AF102507) fizzy-related protein [Homo sapiens]	7E-60
521	3237304	(U91561) pyridoxine 5'-phosphate oxidase [Rattus norvegicus]	6E-37
522	2565196	(AF000381) non-functional folate binding protein [Homo sapiens]	1E-17
523	3108057	(AF060539) channel interacting PDZ domain protein [Mus musculus]	3E-63
524	4160432	(AF071592) kinesin superfamily motor KIF4 [Homo sapiens]	8E-62
525	423916	myosin-I, Myr 1b (alternatively spliced) - rat	1E-66
526	2687591	(AF033201) clipper/cleavage and polyadenylation specificity factor 30 kDa subunit homolog [Mus musculus]	4E-69
527	464555	RAS-RELATED PROTEIN RAB-12 >gi|206531	6E-70
529	2737967	(U82992) envelope glycoprotein [Human immunodeficiency virus type 1]	9.6
530	1351047	SCARLET PROTEIN >gi|1079665	7.9
531	2924445	(AL022022) PE_PGRS [Mycobacterium tuberculosis]	7.5
532	2078307	(U67264) AcMNPV ORF8/ORF1629 homolog [Helicoverpa zea nuclear polyhedrosis virus]	4.5
533	2078307	(U67264) AcMNPV ORF8/ORF1629 homolog [Helicoverpa zea nuclear polyhedrosis virus]	4.4
534	972711	(L47121) bacteriocin [Carnobacterium piscicola]	4.2
535	2895941	(AF047011) prointerleukin-1 alpha [Canis familiaris]	2.5
536	283868	collagen alpha 1(XI) chain - chicken	2.4
537	2052126	(Z94752) hypothetical protein Rv0992c	0.17

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
538	2317926	(U97553) complement regulatory protein [murine herpesvirus 68]	0.0006
539	3242649	(AB015440) alpha 1 type I collagen [Rana catesbeiana]	0.98
540	540952	hypothetical protein - Pseudomonas aeruginosa aeruginosa]	2.6
542	4886288	(AL050300) putative protein [Arabidopsis thaliana]	0.22
543	3322778	(AE001225) conserved hypothetical protein [Treponema pallidum]	9.6
544	1772556	(Y07850) neurofibromin [Hylobates concolor] >gi|1772563|emb|CAA69179| (Y07853) Neurofibromin [Homo sapiens] >gi|1772576|emb|CAA69180|	9.5
545	1083477	protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form D precursor - mouse	0.08
546	1326298	(U58736) Similar to cuticular collagen. [Caenorhabditis elegans]	0.005
547	4007418	(AF071538) Ets transcription factor PDEF [Homo sapiens]	2E-70
548	4504469	homeo box B5 homeotic protein Hox 2.1 - human >gi|184293 (M92299) homeobox protein [Homo sapiens]	0.64
550	3367649	(Y16349) convulxin alpha [Crotalus durissus]	9.7
551	2808634	(AJ001909) transcriptional activator	0.69
553	1127550	(U18939) orf1 [Battrachocottus baikalensis]	4.6
555	2498512	LDLC PROTEIN protein LDLC - human >gi|575654|emb|CAA84427| (Z34975) ldlCp [Homo sapiens]	6.5
556	5579050	(AL096874) hypothetical protein	3.5
557	3327421	(U97068) zonadhesin [Mus musculus]	4.4
558	2493011	PROBABLE CALCIUM-TRANSPORTING ATPASE 8 >gi|1078570|pir||S54520 probable membrane protein YMR162c - yeast (Saccharomyces cerevisiae) cerevisiae]	3.3
559	3242240	(AJ225122) hyperpolarization-activated cation channel, HAC1 [Mus musculus]	1.1
560	780367	(L41686) ORF [Rattus norvegicus]	1.1
561	3327226	(AB014606) KIAA0706 protein [Homo sapiens]	0.41
562	4886501	(AL050275) hypothetical protein [Homo sapiens]	1.1
565	5105067	(AP000061) 111aa long hypothetical protein [Aeropyrum pernix]	0.51
566	1079404	filamin, Mueller cell - chicken >gi|392018	4.2
567	4680264	(AF121977) odorant receptor S25	2.4
570	4927208	(AF133913) ARL-6 interacting protein-6 [Mus musculus]	5E-29
571	1749774	(Y10018) ANON-66Db [Drosophila melanogaster]	0.079
572	5104722	(AP000060) 224aa long hypothetical protein [Aeropyrum pernix]	9.9

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)			
SEQ ID	ACC'N	DESCRIP.	P VALUE
573	320133	carbonate dehydratase (EC 4.2.1.1) - tiger shark (fragments) >gi|226952|prf||1612265A carbonic anhydrase [Galeocerdo cuvier]	5.7
574	1839000	(Z85982) hypothetical protein Rv1648	4.3
575	1839000	(Z85982) hypothetical protein Rv1648	4.2
576	1885385	(U87863) SNAP-25 interacting protein hrs-2 [Rattus norvegicus]	3.2
577	1055150	(U40028) weak similarity to glycoprotein H precursor K04H4.3 and C05B5.5; glycine-rich [Caenorhabditis elegans]	2.5
578	4262315	(AF075256) nonstructural polyprotein	1.1
579	1001674	(D64002) hypothetical protein	0.1
580	2224707	(AB002381) KIAA0383 [Homo sapiens]	0.027
581	765157	(S74099) polyprotein I(p1, p2, p10, p15/PR=protease, p19=matrix protein, p27/CA=capsid protein, p12/NC=nuclear capsid protein) [avian myeloblastosis virus AMV, Peptide, 701 aa] [Avian myeloblastosis virus]	4.3
582	1079438	ribonucleoprotein - chicken >gi|550458 gallus]	0.85
585	4584062	(AJ011380) polyprotein [porcine enterovirus 1]	7.6
586	3874412	(Z70034) similarity to 35.1KD hypothetical yeast protein (Swiss Prot accession number P38805); cDNA EST CEMSE65F comes from this gene; cDNA EST EMBL:T01315 comes from this gene; cDNA EST yk452e10.3 comes from this gene; cDNA ES... >gi|3877079|emb|CAA90124| (Z49910) similarity to 35.1KD hypothetical yeast protein (Swiss Prot accession number P38805); cDNA EST CEMSE65F comes from this gene; cDNA EST EMBL:T01315 comes from this gene; cDNA EST yk452e10.3 comes from this gene; cDNA ES...	0.23
587	3123910	(AF039204) methyltransferase/helicase polyprotein	5.7
589	4539761	(AF118391) salivary peroxidase	3.2
590	115317	COLLAGEN ALPHA 1(VIII) CHAIN PRECURSOR (ENDOTHELIAL COLLAGEN) >gi|89957|pir||A34246 collagen alpha 1(VIII) chain precursor - rabbit	0.02
591	4758548	Homer, neuronal immediate early gene, 2 >gi|3834619 (AF093264) homer-2b [Homo sapiens]	2E-18
592	3219961	PUTATIVE HELICASE C17H9.02 IN CHROMOSOME I >gi|2330709|emb|CAB11211.1| (Z98597) putative helicase [Schizosaccharomyces pombe]	7.3

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
594	5459418	(Y18285) mannose binding lectin-associated serine protease-2 [Rattus norvegicus]	3.3
595	4127783	(AJ130871) Bazooka protein [Drosophila melanogaster]	2.3
596	563601	(X78602) hypothetical replicase [Peanut clump virus]	6.6
599	1778663	(D83674) MesP1 [Mus musculus]	2.4
600	404789	(L22756) GTG start codon; ORFA [Bradyrhizobium japonicum]	0.027
603	1743404	(Z83327) transport-associated protein	3
604	1438537	(U49058) rA4 [Rattus norvegicus]	2
605	3929698	(AL031863) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=68.61; 1-evidence_end; 2-evidence=predicted by match; 2-match_accession=AA541052; 2-match_description=LD20837.5prime LD Drosophila melanogaster...	0.83
606	2708329	(AF038564) atrophin-1 interacting protein 4 [Homo sapiens]	5E-14
608	4099321	(U86145) neuraminidase [influenza A virus	5.8
609	3881475	(Z82083) ZK1010.2 [Caenorhabditis elegans]	4E-12
612	220578	(D00570) open reading frame (251 AA) [Mus musculus]	0.056
613	4826716	equilibrative nucleoside transporter 1 >gi|1845345 (U81375) equilibrative nucleoside transporter 1 [Homo sapiens] >gi|3694940 transporter [Homo sapiens]	0.000008
615	2952333	(AF049885) Arg/Abl-interacting protein ArgBP2b [Homo sapiens]	1.9
618	727264	(U18791) hydroxyproline-rich glycoprotein precursor	4.3
619	4507009	solute carrier family 25 member 14 >gi|3851540 (AF078544) brain mitochondrial carrier protein-1 [Homo sapiens] mitochondrial carrier protein-1 (BMCP1)) [Homo sapiens]	8E-36
620	4884108	(AL050089) hypothetical protein [Homo sapiens]	4E-41
622	113083	ACETYLCHOLINE RECEPTOR PROTEIN, BETA CHAIN PRECURSOR >gi|112056|pir||S13873 nicotinic acetylcholine receptor beta chain precursor - rat beta-subunit [Rattus rattus]	3.3
623	3757569	(AL031863) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]	0.65

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
624	422320	protein kinase (EC 2.7.1.37) - Plasmodium falciparum >gi|9878|emb|CAA47704| (X67288) protein kinase [Plasmodium falciparum] >gi|3845284 (AE001419) calcium-dept. protein kinase (C-term. EF hand)	7.5
625	2496852	HYPOTHETICAL 131.5 KD PROTEIN C02F12.7 IN CHROMOSOME X >gi|1109896 (U41545) coded for by C. elegans cDNA yk4b2.5; coded for by C. elegans cDNA CEESN67F; coded for by C. elegans cDNA yk94h12.5; coded for by C. elegans cDNA CEESD93F; coded for by C. elegans cDNA CEESG57F; coded for by C. elegans cDNA yk4b2.3;...	0.0001
629	1361305	IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog SepA precursor - Shigella flexneri flexneri]	4.2
635	3878636	(Z49128) similar to cAMP-dependant protein kinase; cDNA EST EMBL:T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk492f4.3 comes from this gene; cDNA EST y...	1E-39
636	3649791	(AB012917) serine protease (TLSP) [Homo sapiens]	8E-42
637	868241	(U29488) C56C10.3 gene product [Caenorhabditis elegans]	7E-14
640	2224593	(AB002324) KIAA0326 [Homo sapiens]	4E-25
642	4185794	(AF097025) cysteine desulfurase [Homo sapiens]	1E-64
643	1083755	phosphoprotein phosphatase (EC 3.1.3.16) PPT	2E-15
644	5525067	(AL096844) probable 3-oxacyl-(acyl-carrier-protein) reductase [Streptomyces coelicolor A3(2)]	2E-19
645	4151929	(AF110377) PCAF-associated factor 400 [Homo sapiens]	0.003
646	1174664	RHODOCOXIN >gi|576672 (U17130) ThcC	0.85
648	2702397	(AF038608) Contains similarity to Pfam domain: PF00046 (homeobox), Score=81.5, E-value=5.5e-21, N=1 [Caenorhabditis elegans]	1.6
649	4506891	SET translocation (myeloid leukemia-associated) >gi|346361|pir||A45018 template activating factor-I, splice form beta - human >gi|338039	3E-10
650	4758006	chloride intracellular channel 3 chloride channel CLIC3 [Homo sapiens]	9E-13
651	3702453	(AL021366) cICK0721Q.3 (Kinesin related protein) [Homo sapiens]	5E-38
654	2276316	(Z96810) GLYT-1 LIKE [Homo sapiens]	7E-53
655	3599478	(AF085185) Myosin-IA [Acanthamoeba castellanii]	0.18
657	2735147	(U87971) syntaxin 5 [Rattus norvegicus]	3E-08

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
658	4557651	heat shock transcription factor 4 transcription factor 4 [Homo sapiens]	3E-23
659	4557651	heat shock transcription factor 4 transcription factor 4 [Homo sapiens]	3E-23
661	4505135	midkine (neurite growth-promoting factor 2) >gi|127116|sp|P21741|MK_HUMAN MIDKINE PRECURSOR (NEURITE OUTGROWTH-PROMOTING PROTEIN) (MK) OUTGROWTH-PROMOTING FACTOR 2) >gi|88156|pir||JH0385 midkine precursor - human >gi|35087|emb|CAA38908| sapiens] >gi|182651 (M69148) midkine [Homo sapiens] sapiens] >gi|219929|dbj|BAA01457| (D10604) midkine [Homo sapiens]	2E-15
662	1708021	GLYPICAN-2 PRECURSOR (CEREBROGLYCAN) precursor - rat >gi|440127 (L20468) cerebroglycan cerebroglycan [Rattus norvegicus]	0.00004
663	4504279	H3 histone, family 3A 3B (H3.3B) >gi|122075|sp|P06351|H33_HUMAN HISTONE H3.3 rabbit >gi|90624|pir||S04186 histone H3.3 - mouse histone H3.3 - fruit fly (Drosophila melanogaster) histone H3.3B - chicken >gi|2119023|pir||S61218 histone H3.3 - fruit fly (Drosophila hydei) histone (AA 1-136) [Oryctolagus cuniculus] 136) [Gallus gallus] >gi|161190 (M17876) histone H3 sapiens] >gi|306849 (M11353) H3.3 histone [Homo sapiens] norvegicus] >gi|761716|emb|CAA88778| (Z48950) histone H3.3 [Homo sapiens] >gi|963024|emb|CAA57078| (X81206) histone H3.3 [Drosophila hydei] >gi|963026|emb|CAA57081|	6E-47
664	4557471	coat assembly complex AP1 sigma-1A subunit >gi|231555|sp|Q00382|AP19_MOUSE CLATHRIN COAT ASSEMBLY PROTEIN AP19 (CLATHRIN COAT ASSOCIATED PROTEIN AP19) (GOLGI ADAPTOR AP-1 19 KD ADAPTIN) (HA1 19 KD SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN COMPLEX 1 SMALL CHAIN) >gi|109674|pir||A40535 clathrin-associated protein 19 - mouse >gi|191983 AP-1 clathrin adaptor complex [Homo sapiens]	2E-64
671	410607	drebrin A [chickens, Peptide, 653 aa]	5.4
672	5031433	(AF152396) beta-lactamase-like protein [Mycobacterium fortuitum]	2.3

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
673	1346125	GROWTH/DIFFERENTIATION FACTOR 5 PRECURSOR (GDF-5) (CARTILAGE-DERIVED MORPHOGENETIC PROTEIN 1) (CDMP-1) >gi|1082279|pir||A55452 cartilage-derived morphogenetic protein 1 precursor - human >gi|600732 (U13660) cartilage-derived morphogenetic protein 1 precursor [Homo sapiens]	1.4
674	1730569	PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE TYPE III (1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE KINASE) (PIP5KIII) (PTDINS(4)P-5-KINASE C ISOFORM) 1-phosphatidylinositol-4-phosphate 5-kinase (EC 2.7.1.68) isoform C - human >gi|1042034|bbs|169311 isoform C, PtdIns4P 5-kinase isoform C [human, peripheral blood leukocytes, Peptide, 406 aa] [Homo sapiens]	1.2
675	121781	ENDOGLUCANASE A Bacillus sp >gi|142660 (M14781) cellulase (EC 3.2.1.4)	0.8
676	2497556	PUTATIVE MOLLUSCAN INSULIN-RELATED PEPTIDE(S) RECEPTOR PRECURSOR >gi|1020140|emb|CAA59353| peptide(s) [Lymnaea stagnalis]	0.28
677	1330328	(U50595) Rab8-interacting protein [Mus musculus]	0.096
678	5689505	(AB029007) KIAA1084 protein [Homo sapiens]	4E-59
679	3876327	(Z79754) Similarity to some phosphatases and kinases; cDNA EST EMBL:Z14643 comes from this gene	5E-10
680	4589530	(AB023160) KIAA0943 protein [Homo sapiens]	1E-73
681	533891	(L36073) T-cell receptor antigen [Mus musculus] musculus]	0.31
683	137889	HYPOTHETICAL GENE 3 PROTEIN ictalurid herpesvirus 1 (strain auburn 1) >gi|331213 4886-5794 [Ictalurid herpesvirus 1]	1.6
684	421057	hypothetical protein - Escherichia coli plasmid R100 >gi|42624|emb|CAA39338| (X55815) open reading frame [Escherichia coli]	0.26

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
685	3876064	(Z72507) similar to Thrombospondin type 1 domain; cDNA EST EMBL:D34389 comes from this gene; cDNA EST EMBL:D37437 comes from this gene; cDNA EST EMBL:D64645 comes from this gene; cDNA EST EMBL:D65908 comes from this gene; cDNA ... >gi|3877441|emb|CAA96654| EST EMBL:D34389 comes from this gene; cDNA EST EMBL:D37437 comes from this gene; cDNA EST EMBL:D64645 comes from this gene; cDNA EST EMBL:D65908 comes from this gene; cDNA ...	4.5
686	2317934	(U97553) unknown [murine herpesvirus 68]	0.02
689	1377886	(L46815) DNA binding protein Rc [Mus musculus]	4.7
690	627570	phosphorylation regulatory protein HP-10 - human	1.6
691	480485	cytochrome-c oxidase (EC 1.9.3.1) chain III - Herpetomonas mariadeanei mitochondrion (SGC6)	1.2
692	4885599	SKI-like SNON >gi|68923|pir||TVHUSN transforming protein sno-N - human >gi|36511|emb|CAA33289| (X15219) snoN protein (AA 1 - 684) [Homo sapiens]	0.18
693	3927838	(AC005727) unknown protein [Arabidopsis thaliana]	0.000007
694	5104854	(AP000061) 522aa long hypothetical protein [Aeropyrum pernix]	2.6
698	4240173	(AB020649) KIAA0842 protein [Homo sapiens]	4E-39
699	4096674	(U35833) ARX [Mus musculus]	5E-16
700	117525	LYCOPENE CYCLASE	6.1
701	4049765	(AF063866) ORF MSV249 hypotehtical protein [Melanoplus sanguinipes entomopoxvirus]	8.1
702	4240203	(AB020664) KIAA0857 protein [Homo sapiens]	8E-43
703	3874634	(Z68159) Similarity to Yeast DNA repair protein RAD50 (SW:RA50_YEAST); cDNA EST EMBL:D37313 comes from this gene; cDNA EST EMBL:D34285 comes from this gene [Caenorhabditis elegans]	3.4
704	201995	(M64866) thrombospondin [Mus musculus]	2.3
705	118288	LARIAT DEBRANCHING ENZYME debranching enzyme [Saccharomyces cerevisiae] >gi|172552 cerevisiae] >gi|486256|emb|CAA81990| (Z28149) ORF YKL149c [Saccharomyces cerevisiae]	1.9
706	2654898	(AF016121) envelope protein 2 [Hepatitis GB virus C]	1.6
707	5701582	(AF026205) No definition line found [Caenorhabditis elegans]	1.5
708	2327063	(AF001305) protease 1 [Pneumocystis carinii f. sp. carinii]	0.18
709	422761	basonuclin - human	0.17

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
710	71403	collagen alpha 1(I) chain - rat (fragments)	0.007
714	5105952	(AP000064) 101aa long hypothetical protein [Aeropyrum pernix]	9.7
717	1168479	APX-1 PROTEIN PRECURSOR >gi|473871	6.3
718	4929024	(AF139719) unknown [Klebsiella oxytoca]	0.49
719	417509	GENOME POLYPROTEIN [CONTAINS: NUCLEAR INCLUSION PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) ; COAT PROTEIN (CP)] >gi|320062|pir||GNVSMB genome polyprotein - maize dwarf mosaic virus (strain B) protein [Maize dwarf mosaic virus]	0.51
720	1351287	TROPOMYOSIN 1, FUSION PROTEIN 34 exons [Drosophila melanogaster]	0.11
721	3005601	(AF052433) katanin p80 subunit [Strongylocentrotus purpuratus]	2E-16
722	1360769	DNA helicase-primase complex component - equine herpesvirus 2 >gi|695213 (U20824) DNA helicase-primase complex component [Equine herpesvirus 2]	2
723	5689525	(AB029017) KIAA1094 protein [Homo sapiens]	1E-28
724	3876982	(Z81536) F40D4.11 [Caenorhabditis elegans]	7.7
725	85437	neurofilament triplet M protein - Pacific electric ray (fragment)	0.011
727	2978255	(AB007407) myeloid zinc finger protein-2	0.42
730	4091914	(AF064823) NADH dehydrogenase subunit 5 [Sarcophyton glaucum]	3.5
731	2905612	(AF041845) gp130p1 [Xenopus laevis]	2.7
732	2905612	(AF041845) gp130p1 [Xenopus laevis]	2.7
733	2887499	(AC004143) R29893_1 [Homo sapiens]	2.7
738	4587223	(AB021660) carbonic anhydrase VB [Homo sapiens]	3.3
739	4886445	(AL050269) hypothetical protein [Homo sapiens]	1E-14
740	5102812	(AL079308) putative serine/threonine protein kinase [Streptomyces coelicolor]	1.1
741	4539386	(AL035526) extensin-like protein	0.14
742	2496576	HYPOTHETICAL 32.5 KD PROTEIN Y4AD	7.8
743	3882265	(AB018315) KIAA0772 protein [Homo sapiens]	2E-13
744	3875383	(Z54284) D2085.2 [Caenorhabditis elegans]	0.000003
745	3116122	(AL023287) hypothetical protein	3.8
749	3043716	(AB011168) KIAA0596 protein [Homo sapiens]	0.28
750	3168604	(U88154) proline and glutamic acid rich nuclear protein isoform [Homo sapiens]	0.035
752	2429324	(AF015116) interleukin 6 receptor [Sus scrofa]	1.3

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
753	808667	(M15972) The first atg start codon is the AA before the stop codon in ORF1; putative [Human herpesvirus 4]	1.1
754	164840	(M10412) carbonic anhydrase I [Oryctolagus cuniculus]	0.88
755	2133726	synapse-associated protein sap47-1 - fruit fly (Drosophila melanogaster) >gi|929571|emb|CAA56416| melanogaster]	7E-22
756	136797	HYPOTHETICAL PROTEIN UL7 cytomegalovirus (strain AD169) >gi|59612|emb|CAA35440|	7.4
757	3881372	(Z81141) ZC47.14 [Caenorhabditis elegans]	3.3
759	3327160	(AB014573) KIAA0673 protein [Homo sapiens]	3E-57
761	1651655	(D90899) PNIL34 [Synechocystis sp.]	6.3
762	94228	env polyprotein - feline immunodeficiency virus >gi|59290|emb|CAA40321| (X57002) ENV [Feline immunodeficiency virus] >gi|228554|prf||1805419A envelope glycoprotein [Feline immunodeficiency virus]	8.2
763	5102812	(AL079308) putative serine/threonine protein kinase [Streptomyces coelicolor]	0.94
764	5454104	transcriptional adaptor 2 complex) >gi|3335555 (AF069733) ADA3-like protein [Homo sapiens]	2E-54
766	1280102	(U55370) coded for by C. elegans cDNA CEESD82F; coded for by C. elegans cDNA CEESD82R [Caenorhabditis elegans]	4.5
768	3875720	(Z50857) M79.2 [Caenorhabditis elegans] elegans]	4.9
769	4502247	armadillo repeat protein sapiens]	4.8
770	3860231	(AF102887) thrombospondin-4 [Mus musculus]	3.6
771	539999	receptor tyrosine kinase c-kit - rat tyrosine kinase [Rattus rattus]	2.9
773	3550638	(AJ006986) repeating unit transporter	6.5
775	5105066	(AP000061) 124aa long hypothetical protein [Aeropyrum pernix]	7.7
776	2622679	(AE000916) tungsten formylmethanofuran dehydrogenase, subunit A [Methanobacterium thermoautotrophicum]	4.8
777	1086650	(U41015) Similar to serine/threonine protein kinase.	0.4
778	1363837	probable finger protein YOL054w - yeast cerevisiae] >gi|1419863|emb|CAA99062| (Z74796) ORF YOL054w [Saccharomyces cerevisiae]	0.14
779	500858	(D14168) 50kDa lectin [Bombyx mori]	0.0000004
780	4680659	(AF132944) CGI-10 protein [Homo sapiens]	4E-67
785	4586844	(AB015633) type II membrane protein	3E-09
786	117800	CYANAMIDE HYDRATASE (UREA HYDRO-LYASE) >gi|102020|pir||A39365 cyanamide hydratase verrucaria]	1.8

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
787	5689513	(AB029011) KIAA1088 protein [Homo sapiens]	3E-09
790	2829815	CARBON STARVATION PROTEIN A HOMOLOG tuberculosis]	6.9
791	2224671	(AB002363) KIAA0365 [Homo sapiens]	6.6
792	3834629	(AF094519) diaphanous-related formin; p134 mDia2 [Mus musculus]	1E-23
793	1572522	(U67194) upf54.8 [Enterobacter aerogenes]	3.3
794	3876099	(Z75536) similar to dynein heavy chain; cDNA EST EMBL:D27549 comes from this gene; cDNA EST EMBL:D34859 comes from this gene [Caenorhabditis elegans]	0.00001
797	3319990	(Y17267) ubiquitin-conjugating enzyme [Mus musculus]	4E-40
799	473513	(M17619) NADH dehydrogenase subunit COIII [Asterina pectinifera]	2.8
800	1460094	(L35031) Orf159; Predicted integral membrane protein with 4 transmembrane sequences (method of Klein, Kanehisa, DeLisi in PCGene). One nucleotide overlap with upstream orf.; putative [Escherichia coli]	1.6
801	4455041	(AF116463) unknown [Streptomyces lincolnensis]	0.081
802	1174467	STAR PROTEIN >gi|472815 (L31886) amino acid feature: potential transmembrane domain, aa 280 .. 302 [Drosophila melanogaster]	0.053
805	5031861	candidate tumor suppressor involved in B-CLL >gi|3133092|emb|CAA12136| (AJ224819) tumor suppressor [Homo sapiens]	3E-15
807	1947168	(AF000299) No definition line found [Caenorhabditis elegans]	0.24
808	5442104	(AF126467) Gag protein [Simian retrovirus SRV-2]	7.8
809	1684987	(U20649) NADH dehydrogenase subunit [Cymbidium atropurpureum]	6
810	1709814	PHOTOSYSTEM I P700 CHLOROPHYLL A APOPROTEIN A1 >gi|2147916|pir||S73205 photosystem I p700 chlorophyll A apoprotein A1 - Porphyra purpurea chloroplast >gi|1276750 (U38804) Photosystem I p700 chlorophyll A apoprotein A1 [Porphyra purpurea]	0.74
811	400280	MELANOCYTE STIMULATING HORMONE RECEPTOR (MSH-R) (MELANOTROPIN RECEPTOR) (MELANOCORTIN-1 RECEPTOR) (MC1-R) >gi|110690|pir||S25581 melanocyte-stimulating hormone receptor - mouse hormone receptor [Mus musculus]	10

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
812	3581887	(AL031540) internalin- related, Leucine rich repeat containing protein [Schizosaccharomyces pombe]	3.6
813	2462671	(Z98529) putative RNA-binding protein	0.002
814	2492678	ACTIN-LIKE PROTEIN ARP8 YOR141c - yeast (Saccharomyces cerevisiae)	1E-15
815	3879157	(Z77668) predicted using Genefinder; Similarity to Mouse selenium-binding protein	6
816	3668141	(AJ007398) PBK1 protein [Homo sapiens]	8E-57
817	3875131	(Z70750) similar to vanadate resistance protein transmembranous domains [Caenorhabditis elegans]	5E-33
819	2494509	PUTATIVE FORKHEAD-RELATED TRANSCRIPTION FACTOR F26A1.2 >gi|860690 (U27312) weak similarity to FKH-5 Protein (Mouse, PIR:S36074) and D. melanogaster fork head domain protein FD4	9.3
820	547625	TRANSCRIPTION FACTOR HES-1 (HAIRY AND ENHANCER OF SPLIT 1) >gi|539928|pir||A53336 transcription factor HES-1 - mouse factor HES-1 [Mus musculus]	0.69
821	113670	!!!! ALU CLASS E WARNING ENTRY !!!!	0.23
822	3329124	(AE001337) S/T Protein Kinase [Chlamydia trachomatis]	2.5
823	4982299	(AE001811) conserved hypothetical protein [Thermotoga maritima]	0.09
824	4530509	(AF124748) putative RNA-binding protein	3
825	75198	glycoprotein precursor - Uukuniemi virus	0.59
826	127477	MEMBRANE-ASSOCIATED ATPASE GAMMA CHAIN (SUL-ATPASE GAMMA) (ATP SYNTHASE, SUBUNIT D) 3.6.1.34) gamma chain - Sulfolobus acidocaldarius	0.2
827	3915729	HYPERPLASTIC DISCS PROTEIN (HYD PROTEIN) >gi|2673887 (L14644) hyperplastic discs protein	0.22
836	4493951	(AL034556) predicted using hexExon; MAL3P5.16 (PFC0650w), Hypothetical protein, len: 1282 aa	0.69
837	4884027	(AJ011655) hypothetical protein	2.5
838	3873691	(Z46240) similar to endothelial actin-binding protein repeats; cDNA EST EMBL:D27639 comes from this gene; cDNA EST EMBL:D33624 comes from this gene; cDNA EST EMBL:D33507 comes from this gene; cDNA EST EMBL:D36493 comes from thi...	9.7

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
839	4507425	teratocarcinoma-derived growth factor 1 >gi|117473|sp|P13385|CRI1_HUMAN TERATOCARCINOMA-DERIVED GROWTH FACTOR 1 (EPIDERMAL GROWTH FACTOR-LIKE CRIPTO PROTEIN CR1) (CRIPTO-1 GROWTH FACTOR) (CRGF) >gi|87385|pir||A30362 epidermal growth factor-like protein CR3 - human >gi|30221|emb|CAA32467| factor 1 [Homo sapiens]	0.49
840	2271518	(AF009829) unknown [Mycobacterium bovis]	0.082
841	4093025	(AF070836) NADH dehydrogenase subunit 4	1.5
842	2662603	(AF036699) No definition line found	6.4
843	3880368	(Z95621) similar to collagen; cDNA EST EMBL:D69870 comes from this gene; cDNA EST EMBL:D70498 comes from this gene [Caenorhabditis elegans] cDNA EST EMBL:D69870 comes from this gene; cDNA EST EMBL:D70498 comes from this gene [Caenorhabditis elegans]	3.3
844	731490	HYPOTHETICAL 73.0 KD PROTEIN IN SEB1-PTC2 INTERGENIC REGION >gi|1077682|pir||S50591 hypothetical protein YER088c - yeast (Saccharomyces cerevisiae) >gi|603326 (U18839) Yer088cp [Saccharomyces cerevisiae]	1.7
852	4929605	(AF151826) CGI-68 protein [Homo sapiens]	1E-61
855	4996369	(AB021267) polyprotein [Arabidopsis thaliana]	2.2
856	4100563	(AF001175) ribonuclease P protein subunit p14 [Homo sapiens]	2E-10
863	458692	(U06631) homologous to mouse gene PC326:GenBank Accession Number M95564 [Homo sapiens]	6
864	123518	RNA POLYMERASE PRINCIPAL SIGMA FACTOR HRDA >gi|80717|pir||S17929 transcription initiation factor sigma hrdA - Streptomyces coelicolor subunit (AA 1-396) [Streptomyces coelicolor]	3.1
865	126296	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG protein [Nycticebus coucang]	0.0001
866	4972730	(AF132172) unknown [Drosophila melanogaster]	3E-19
867	128169	HIGH-MOLECULAR WEIGHT COBALT-CONTAINING NITRILE HYDRATASE SUBUNIT ALPHA hydratase (EC 4.2.1.84) - Rhodococcus rhodochrous rhodochrous]	5
868	2809262	(AC002560) F21B7.31 [Arabidopsis thaliana]	1.9

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
869	4758374	four and a half LIM domains 3 LIM-protein FHL3 [Homo sapiens]	0.94
871	400784	ACIDIC PHOSPHOPROTEIN PRECURSOR (50 KD ANTIGEN) >gi|477254|pir||A48455 acidic phosphoprotein PcEMA1q - Plasmodium chabaudi >gi|160603 (M95789) acidic phosphoprotein [Plasmodium chabaudi]	3.3
872	2495704	HYPOTHETICAL PROTEIN KIAA0129 product is novel. [Homo sapiens]	0.0002
874	4514345	(AB013374) Ykok [Bacillus halodurans]	3.7
875	5453794	nucleolar protein (KKE/D repeat)	1E-18
877	2650666	(AE001107) A. fulgidus predicted coding region AF2427	0.076
878	1572836	(U70858) similar to family 18 of glycosyl hydrolases	2.7
879	400853	VITAMIN-K DEPENDENT PROTEIN C PRECURSOR (AUTOPROTHROMBIN IIA) (ANTICOAGULANT PROTEIN C) >gi|112216|pir||S18994 protein C (activated) (EC 3.4.21.69) precursor - rat >gi|56963|emb|CAA45617|	4.7
880	113668	!!!! ALU CLASS C WARNING ENTRY !!!!	4.9
881	5262748	(AJ133120) Proline rich synapse associated protein 2 [Rattus norvegicus]	8.6
883	1352130	CYTOCHROME C OXIDASE POLYPEPTIDE I subunit I [Chondrus crispus]	9.9
886	547708	TRANSCRIPTIONAL REGULATOR IE63 human herpesvirus 1 (strain HFEM)	0.31
887	4758216	dishevelled 2 (homologous to Drosophila dsh) >gi|2291008|gb|AAB65243.1| (AF006012) dishevelled 2 [Homo sapiens]	0.051
888	2291257	(AF016430) contains similarity to a BR-C/TTK domain	0.016
889	2911858	(AF047659) No definition line found [Caenorhabditis elegans]	3E-26
890	1932813	(U88065) dsRNA adenosine deaminase [Xenopus laevis]	3.4
893	728836	!!!! ALU SUBFAMILY SP WARNING ENTRY	0.39
895	2580578	(AF000996) ubiquitous TPR motif, Y isoform [Homo sapiens]	0.008
896	1869831	(Z86099) UL9 [human herpesvirus 2]	9.9
897	2632151	(Y14493) PHOX2b protein [Mus musculus] musculus]	2.6
898	1079078	GCR 101 protein - fruit fly (Drosophila melanogaster) >gi|510509|emb|CAA50795| (X71975) put. homologue to S.cerevisiae GAR1 gene [Drosophila melanogaster]	0.0000004
900	418745	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Crithidia oncopelti mitochondrion (SGC6) subunit 4 [Crithidia oncopelti]	4.6

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
901	3874002	(Z36719) cDNA EST yk208g3.5 comes from this gene [Caenorhabditis elegans]	0.41
903	731630	HYPOTHETICAL 50.6 KD PROTEIN IN RPL14B-GPA1 INTERGENIC REGION >gi|626594|pir||S46802 hypothetical protein YHR004c - yeast (Saccharomyces cerevisiae) >gi|500822 (U10555) Yhr004cp [Saccharomyces cerevisiae]	0.42
904	5081459	(AF124435) p55-related MAGUK protein DLG3 [Danio rerio]	6E-27
906	2497012	HYPOTHETICAL 26.6 KD PROTEIN T19C3.4 IN CHROMOSOME III >gi|849238 (U28412) similar to polyposis locus protein 1 (SP:DP1_HUMAN, Q00765)	3E-34
907	113671	!!!! ALU CLASS F WARNING ENTRY !!!!	0.0002
909	5410448	(AF135183) Recq helicase 5	2E-52
910	1407655	(U58884) SH3P7 [Mus musculus]	0.000002
912	2981631	(AB012223) ORF2 [Canis familiaris]	0.000002
913	2289030	(U53564) N-terminal region of the protein [Mus musculus]	3.6
914	4567179	(AC007228) BC37295_1 [Homo sapiens]	0.000002
915	2842531	(AB004291) gamma-subunit of enolase	4.4
916	1177607	(X92485) pva1 [Plasmodium vivax]	0.33
918	3638957	(AC004877) sco-spondin-mucin-like; similar to P98167 uncertain [Homo sapiens]	4.3
919	5032163	transcription factor 17	1E-23
920	3873738	(Z37983) contains five copies of the EGF-like aspartic acid and asparagine hydroxylation site comes from this gene; cDNA EST EMBL:D27753 comes from this gene; cDNA EST ...	4.6
921	4503561	epithelial membrane protein 2 PROTEIN-2 (EMP-2) (XMP PROTEIN) >gi|2474096 (U52100) XMP	4E-08
922	4506051	primase, polypeptide 1 (49kD) SUBUNIT (DNA PRIMASE 49 KD SUBUNIT) (P49) p48) [Homo sapiens]	0.064
923	2738451	(AF003534) putative tyrosine protein kinase [Chilo iridescent virus]	5E-08
925	543222	glutamine (Q)-rich factor 1, QRF-1 - mouse factor 1, QRF-1 [mice, B-cell leukemia, BCL1, Peptide Partial, 84 aa]	2E-44
927	961466	(D63777) adhesive plaque matrix protein	4.9

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
928	121404	ANAEROBIC GLYCEROL-3-PHOSPHATE DEHYDROGENASE SUBUNIT B (G-3-P DEHYDROGENASE) subunit B [Escherichia coli] >gi|1788575 (AE000314) sn-glycerol-3-phosphate dehydrogenase (anaerobic), membrane anchor subunit [Escherichia coli] dehydrogenase (EC 1.1.99.5) (anaerobic) chain B	0.49
930	5640009	(AF167316) zinc finger protein ZFP109 [Mus musculus]	1.2
931	4678836	(AL049701) hypothetical protein [Homo sapiens]	8E-43
932	4758556	U4/U6-associated RNA splicing factor >gi|2708307 (AF016370) U4/U6 small nuclear ribonucleoprotein hPrp3 [Homo sapiens]	6E-09
933	729079	CARBOXY-CIS,CIS-MUCONATE CYCLASE 3-carboxy-cis,cis-muconate cyclase [Neurospora crassa]	6.4
934	5080758	(AC007842) BC331191_1 [Homo sapiens]	2E-08
936	4506009	protein phosphatase 1, regulatory subunit 10 >gi|2117159|emb|CAA73697| (Y13247) FB19 protein [Homo sapiens]	8E-32
937	726403	(U23175) similar to anion exchange protein [Caenorhabditis elegans]	1E-25
938	726403	(U23175) similar to anion exchange protein [Caenorhabditis elegans]	3E-26
940	2500573	RIBONUCLEASE S-4 PRECURSOR (STYLAR GLYCOPROTEIN 4) (S4-RNASE) >gi|1405426|emb|CAA65320|	3.2
941	2291171	(AF016420) No definition line found [Caenorhabditis elegans]	8.7
942	3914191	UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 110 KD SUBUNIT (O-GLCNAC TRANSFERASE P110 SUBUNIT) >gi|1931579 (U76557) O-GlcNAc transferase, p110 subunit [Rattus norvegicus]	1E-17
943	549341	MAJOR CAPSID PROTEIN L1 type 34 >gi|396996|emb|CAA52560| (X74476) late protein	8.1
944	4680673	(AF132951) CGI-17 protein [Homo sapiens]	3E-65
945	4191610	(AF117107) IGF-II mRNA-binding protein 2 [Homo sapiens]	1E-51
947	4589512	(AB023151) KIAA0934 protein [Homo sapiens]	5E-46
948	2193870	(D84391) reverse transcriptase [Mus musculus]	2E-09
949	3046871	(AB003753) high sulfur protein B2E [Rattus norvegicus]	5.7

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
950	140130	HYPOTHETICAL 85.7 KD PROTEIN (ORF C-792) >gi|76733|pir||S03232 hypothetical protein C-792	8
951	2494897	PERIODIC TRYPTOPHAN PROTEIN 1 HOMOLOG (KERATINOCYTE PROTEIN IEF SSP 9502) >gi|177765 sapiens]	2E-08
954	2465332	(U92819) unnamed HERV-H protein [Homo sapiens]	8E-14
955	4200446	(AF102777) FYVE finger-containing phosphoinositide kinase [Mus musculus]	8E-15
956	3859560	(AF098668) acyl-protein thioesterase [Homo sapiens]	2E-61
958	4008551	(AL034490) pseudouridylate synthase	6.7
959	4508041	zinc finger protein 91 (HPF7, HTF10) >gi|549839|sp|Q05481|ZN91_HUMAN ZINC FINGER PROTEIN 91 (ZINC FINGER PROTEIN HTF10) (HPF7)	4E-19
961	1085397	taurine transporter - human >gi|559853	2E-14
962	134211	SERUM RESPONSE FACTOR ACCESSORY PROTEIN 1B (SAP-1B) (ETS-DOMAIN PROTEIN ELK-4) protein-1 form b, SAP-1b - human >gi|338035 (M85164) SAP-1B protein [Homo sapiens]	0.038
963	961444	(D63876) KIAA0154 gene product is related to mouse gamma adaptin. [Homo sapiens]	1E-20
964	4758488	general transcription factor IIF, polypeptide 2 (30kD subunit) FACTOR IIF, BETA SUBUNIT (TFIIF-BETA) (TRANSCRIPTION INITIATION FACTOR RAP30) >gi|105393|pir||S18677 ATP-dependent RNA helicase RAP30/74 chain RAP30 - human RAP30 [Homo sapiens]	0.00009
971	5262560	(AL080125) hypothetical protein [Homo sapiens]	2E-41
972	1707274	(U80931) strong similarity to class-III of pyridoxal-phoshate-dependent aminotransferases	7E-31
973	3810839	(AL032684) conserved hypothetical zinc-finger protein [Schizosaccharomyces pombe]	7E-12
975	4887229	(AF150755) microtubule-actin crosslinking factor [Mus musculus]	5E-22
976	1675222	(U67203) ACF7 neural isoform 1 [Mus musculus]	4E-22
977	987050	(X65335) lacZ [Cloning vector pSV-beta-Galactosidase Control]	3E-15
978	5052075	(AF074331) PAPS synthetase-2	8E-63
979	3983573	(AC004839) similar to IgD B-cell receptor-associated protein (BAP); similar to S46997 (PID:g1085495) [Homo sapiens]	8E-58

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
980	3108057	(AF060539) channel interacting PDZ domain protein [Mus musculus]	1E-63
981	1755049	(U55042) myosin X [Bos taurus]	5E-55
982	1783123	(AB000170) endopeptidase 24.16 type M3 endopeptidase 24.16 type M3 [Sus scrofa] type M3 [Sus scrofa] >gi|1783130|dbj|BAA19065| type M3 [Sus scrofa] >gi|1783134|dbj|BAA19067| type M3 [Sus scrofa]	2E-58
983	1098627	(U31079) 47 kDa heat shock protein [Danio rerio]	4.5
984	3649791	(AB012917) serine protease (TLSP) [Homo sapiens]	6E-76
985	1177607	(X92485) pva1 [Plasmodium vivax]	0.13
988	4106444	(AF085692) multidrug resistance-associated protein 3B	0.97
989	3757569	(AL031863) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]	2.5
990	115317	COLLAGEN ALPHA 1(VIII) CHAIN PRECURSOR (ENDOTHELIAL COLLAGEN) >gi|89957|pir||A34246 collagen alpha 1(VIII) chain precursor - rabbit	0.29
991	1743404	(Z83327) transport-associated protein	6.2
992	3878636	(Z49128) similar to cAMP-dependant protein kinase; cDNA EST EMBL:T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk492f4.3 comes from this gene; cDNA EST y...	3E-53
993	728837	!!!! ALU SUBFAMILY SQ WARNING ENTRY	4
994	1098627	(U31079) 47 kDa heat shock protein [Danio rerio]	4.5
995	3649791	(AB012917) serine protease (TLSP) [Homo sapiens]	6E-76
996	115317	COLLAGEN ALPHA 1(VIII) CHAIN PRECURSOR (ENDOTHELIAL COLLAGEN) >gi|89957|pir||A34246 collagen alpha 1(VIII) chain precursor - rabbit	0.29
997	113668	!!!! ALU CLASS C WARNING ENTRY !!!!	0.012
998	3722229	(AF058790) SynGAP-b [Rattus norvegicus]	3.4
999	3876327	(Z79754) Similarity to some phosphatases and kinases; cDNA EST EMBL:Z14643 comes from this gene	6E-33
1000	4886288	(AL050300) putative protein [Arabidopsis thaliana]	0.22
1001	4589530	(AB023160) KIAA0943 protein [Homo sapiens]	1E-73
1003	4063766	(D87895) chitinase [Emericella nidulans]	0.016
1004	4505727	peroxisomal biogenesis factor 3 PROTEIN PEX3 (PEROXIN-3) >gi|3336882|emb|CAA04879| sapiens] >gi|4218426|emb|CAA08904| (AJ009866) Pex3p	e-126
1005	2832671	(AL021712) hypothetical protein	1.7

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1006	1175815	HYPOTHETICAL PROTEIN HI1476 Haemophilus influenzae (strain Rd KW20) >gi|1574317 influenzae Rd]	7.1
1007	4927208	(AF133913) ARL-6 interacting protein-6 [Mus musculus]	4E-28
1008	2327063	(AF001305) protease 1 [Pneumocystis carinii f. sp. carinii]	0.18
1009	2384956	(AF022985) No definition line found [Caenorhabditis elegans]	3E-19
1010	3877495	(Z48583) cDNA EST EMBL:T00483 comes from this gene; cDNA EST EMBL:D64526 comes from this gene; cDNA EST EMBL:D65147 comes from this gene; cDNA EST EMBL:D68484 comes from this gene; cDNA EST EMBL:D67548 comes from this gene; c... >gi|3879229|emb|CAA88749| EST EMBL:D64526 comes from this gene; cDNA EST EMBL:D65147 comes from this gene; cDNA EST EMBL:D68484 comes from this gene; cDNA EST EMBL:D67548 comes from this gene; cDN...	6.7
1011	3355308	(AJ009695) wall-associated kinase 4	0.74
1012	297922	(X66052) D-lactate dehydrogenase	2
1013	4691726	(AF124490) ARF GTPase-activating protein GIT1 [Homo sapiens]	4E-68
1014	2384956	(AF022985) No definition line found [Caenorhabditis elegans]	3E-19
1016	1469880	(D63483) The KIAA0149 gene product is related to Notch3. [Homo sapiens]	0.58
1017	5678967	(AL109630) BACR7A4.ai [Drosophila melanogaster]	1.4
1018	3023956	VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1 >gi|607003 (L28125) beta transducin-like protein	5E-28
1019	3882321	(AB018343) KIAA0800 protein [Homo sapiens]	e-105
1020	1655907	(U65891) protein tyrosine phosphatase CRYP-2 [Gallus gallus]	2.5
1021	3540219	(D87686) KIAA0017 protein [Homo sapiens]	8E-70
1022	1352368	ENTEROPEPTIDASE PRECURSOR enterokinase [Bos taurus]	7.7
1023	4506701	ribosomal protein S23 S23 >gi|543449|pir||S41955 ribosomal protein S23 - rat protein [Homo sapiens] >gi|453281|emb|CAA54584| (X77398) ribosomal protein S23 [Rattus norvegicus]	9E-15
1024	5059323	(AF151522) hairy and enhancer of split related-1 [Homo sapiens]	0.31
1025	3329139	(AE001339) ABC Transporter Membrane Protein [Chlamydia trachomatis]	1.2

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1027	2133726	synapse-associated protein sap47-1 - fruit fly (Drosophila melanogaster) >gi|929571|emb|CAA56416| melanogaster]	8E-25
1028	4507009	solute carrier family 25 member 14 >gi|3851540 (AF078544) brain mitochondrial carrier protein-1 [Homo sapiens] >gi|4678718|emb|CAB41251.1| protein-1 (BMCP1)) [Homo sapiens]	e-121
1029	2662336	(D55702) ORF2 [Bombyx mori]	8.2
1030	5454104	transcriptional adaptor 2 complex) >gi|3335555 (AF069733) ADA3-like protein [Homo sapiens]	e-108
1031	1363044	mucin (clone pGM7-1) - bovine repeats, clone pGBM7-1} [cattle, gall-bladder, Peptide Partial, 600 aa] [Bos taurus]	0.21
1032	3874427	(Z78416) predicted using Genefinder; Similarity to S.pombe RAD18 gene (TR:E198069); cDNA EST CEESX52R comes from this gene; cDNA EST EMBL:D32785 comes from this gene; cDNA EST EMBL:D35528 comes from this gene; cDNA EST EMBL:D37...	6E-09
1033	3287732	GLYCYL-GLYCINE ENDOPEPTIDASE ALE-1 PRECURSOR >gi|1890068|dbj|BAA13069| (D86328) ALE-1	1.7
1034	2143767	glycoprotein - rat >gi|986943 (L08134) glycoprotein [Rattus norvegicus] norvegicus]	0.057
1035	4929167	(AF142440) BC1 [Indian mungbean yellow mosaic geminivirus]	0.63
1036	2224593	(AB002324) KIAA0326 [Homo sapiens]	2E-41
1037	3820909	(AJ010642) Dof protein [Drosophila melanogaster]	1.9
1038	4586844	(AB015633) type II membrane protein	3E-09
1039	3287688	(AC003979) Contains similarity to ycf37 gene product gb|1001425 from Synechocystis sp. genome gb|D63999. ESTs gb|T43026, gb|R64902, gb|Z18169 and gb|N37374 come from this gene. [Arabidopsis thaliana]	0.036
1040	4557651	heat shock transcription factor 4 transcription factor 4 [Homo sapiens]	3E-23
1041	4557651	heat shock transcription factor 4 transcription factor 4 [Homo sapiens]	3E-23

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
1042	4557471	coat assembly complex AP1 sigma-1A subunit >gi|231555|sp|Q00382|AP19_MOUSE CLATHRIN COAT ASSEMBLY PROTEIN AP19 (CLATHRIN COAT ASSOCIATED PROTEIN AP19) (GOLGI ADAPTOR AP-1 19 KD ADAPTIN) (HA1 19 KD SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN COMPLEX 1 SMALL CHAIN) >gi|109674|pir||A40535 clathrin-associated protein 19 - mouse >gi|191983 AP-1 clathrin adaptor complex [Homo sapiens]	4E-73
1043	3581887	(AL031540) internalin- related, Leucine rich repeat containing protein [Schizosaccharomyces pombe]	3.6
1044	2492678	ACTIN-LIKE PROTEIN ARP8 YOR141c - yeast (Saccharomyces cerevisiae)	6E-21
1052	2495704	HYPOTHETICAL PROTEIN KIAA0129 product is novel. [Homo sapiens]	0.0002
1054	4680703	(AF132966) CGI-32 protein [Homo sapiens]	1E-91
1055	1078718	reverse transcriptase - Trypanosoma cruzi transcriptase [Trypanosoma cruzi]	1.1
1056	728835	!!!! ALU SUBFAMILY SC WARNING ENTRY	0.16
1057	1932813	(U88065) dsRNA adenosine deaminase [Xenopus laevis]	5.4
1058	1363325	RNA helicase HEL117 - rat >gi|897915 (U25746) RNA helicase [Rattus norvegicus]	3E-91
1059	631772	TEG-261 protein - mouse	2E-47
1062	2506774	KERATIN, TYPE II CYTOSKELETAL 8	2E-42
1063	4406551	(AF131739) Unknown [Homo sapiens]	2E-82
1064	4678836	(AL049701) hypothetical protein [Homo sapiens]	3E-13
1065	4567179	(AC007228) BC37295_1 [Homo sapiens]	0.000005
1066	3953593	(AB020542) Zinc finger protein s11-6 [Mus musculus]	1E-32
1068	3170180	(AF039690) antigen NY-CO-8 [Homo sapiens]	0.26
1069	4678836	(AL049701) hypothetical protein [Homo sapiens]	3E-13
1070	2558501	(D63850) hepatoma-derived growth factor	3E-24
1071	3327220	(AB014603) KIAA0703 protein [Homo sapiens]	5E-53
1072	5453601	cartilage-associated protein cartilage-associated protein (CASP) [Homo sapiens]	e-125
1073	3237304	(U91561) pyridoxine 5'-phosphate oxidase [Rattus norvegicus]	3E-56
1074	4508041	zinc finger protein 91 (HPF7, HTF10) >gi|549839|sp|Q05481|ZN91_HUMAN ZINC FINGER PROTEIN 91 (ZINC FINGER PROTEIN HTF10) (HPF7)	1E-21

Table 2B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)			
SEQ ID	ACC'N	DESCRIP.	P VALUE
1075	961444	(D63876) KIAA0154 gene product is related to mouse gamma adaptin. [Homo sapiens]	1E-20
1077	1707274	(U80931) strong similarity to class-III of pyridoxal-phoshate-dependent aminotransferases	7E-31
1078	1546779	(U28789) PACT [Mus musculus]	0
1079	5052075	(AF074331) PAPS synthetase-2	8E-63

Table 3

SEQ ID NO:	Profilename	Start	Stop	Direction
97	Kazal	25	243	for
227	helicase_C	212	389	for
242	EFhand	275	310	for
450	SH3	44	226	for
473	Zincfing_C2H2	211	273	for
505	WD_domain	80	178	for
512	Zincfing_C2H2	147	209	for
523	PDZ	168	395	for
527	ras	18	395	for
545	ANK	311	393	for
547	Ets_Nterm	7	237	for
606	WW_domain	120	209	for
635	protkinase	47	400	for
635	mkk	41	394	for
636	trypsin	147	381	for
640	Zincfing_C2H2	122	184	for
693	Zincfing_CCHC	135	185	for
721	WD_domain	18	116	for
805	Zincfing_C3HC4	263	406	for
918	BZIP	51	224	for
919	Zincfing_C2H2	125	187	for
925	FKH	9	230	for
971	Zincfing_C2H2	202	264	for
973	Zincfing_CCHC	262	309	for
980	PDZ	241	468	for
992	mkk	0	708	for
992	protkinase	121	711	for
995	trypsin	202	760	for
984	trypsin	202	760	for
1018	WD_domain	18	116	for
1028	pr55	24	1293	for
1035	ATPases	74	616	for
1036	Zincfing_C2H2	122	184	for
1053	14_3_3	63	619	for
1058	helicase_C	212	448	for
1058	ATPases	59	442	for
1063	Zincfing_C2H2	211	273	for
1066	Zincfing_C2H2	125	187	for
1072	ATPases	808	1284	for
1078	protkinase	309	1022	rev
1078	neur_chan	12	508	rev
1078	Zincfing_CCHC	262	309	for
1078	Zincfing_C3HC4	557	679	for

Table 13			
ES55	**ES56**	**ES57**	**ES58**
M00004170C:H06	M00004036B:C11	M00004288D:E07	M00023298B:G07
M00004170D:C06	M00004064B:G03	M00004318D:D07	M00026819B:E02
M00004171D:H10	M00004067C:E05	M00004356C:D02	M00026914C:H10
M00004174B:B12	M00004099C:F04	M00004391C:F12	M00027023B:H12
M00004175D:G10	M00004103A:E06	M00004386C:C03	M00027085A:G10
M00004176A:E07	M00004128B:H11	M00004414D:C11	M00027248D:D01
M00001352D:A09	M00004167A:H04	M00004422C:A01	M00027546B:A11
M00001345C:B10	M00004158C:B01	M00004427D:H04	M00023299B:A01
M00001382D:F03	M00004165B:E03	M00004502B:G05	M00026857A:F02
M00001419A:E01	M00004181A:B05	M00004495D:A05	M00026858C:H05
M00001437D:A12	M00003993C:G11	M00005364C:A02	M00026861A:B05
M00001441D:G02	M00004046C:A04	M00005375B:H03	M00026846C:B01
M00001601D:A03	M00004034A:G03	M00005420C:E10	M00027131A:H02
M00001677B:G01	M00004036C:E10	M00005413B:B02	M00027396A:F07
M00001678A:B10	M00004043C:A06	M00005438D:A08	M00023301B:C01
M00001675C:F05	M00004067C:C10	M00005453B:B06	M00023321B:F06
M00001360D:C12	M00004068A:A03	M00005446B:D10	M00023401C:D12
M00001389C:E01	M00004069A:E04	M00005493D:H12	M00026941C:E11
M00001390C:H05	M00004071C:B06	M00005476D:A11	M00027067A:B02
M00001399B:C04	M00004127C:C08	M00005482A:D08	M00027036B:D07
M00001507A:H06	M00004157C:E06	M00005485C:F09	M00027329A:H04
M00003747C:G12	M00004165D:H12	M00005563C:D05	M00027740C:C05
M00001358B:F12	M00003995B:C06	M00005569B:E04	M00023340A:A10
M00001360B:F09	M00004090A:B11	M00005621B:C09	M00026942C:A06
M00001392A:F02	M00004084C:F05	M00005628D:A10	M00027066A:A04
M00001397D:G04	M00004087A:H06	M00005629B:G06	M00027072C:A11
M00001463C:E12	M00004110A:G03	M00004866C:H08	M00027028A:B06
M00001531B:A03	M00004117D:F06	M00004872C:G03	M00023282B:H09
M00001507D:F09	M00004150A:B09	M00005358B:D10	M00023295B:C03
M00001513B:F05	M00004140C:D04	M00005385D:B08	M00026811A:H01
M00001514B:C02	M00004175D:D05	M00005392C:B03	M00026850B:F07
M00001576C:E03	M00004176A:H05	M00005395C:C11	M00026913D:G11
M00003756D:B09	M00004170C:A12	M00005396A:C01	M00026936D:D01
M00003907C:D02	M00004237B:G01	M00005435B:F01	M00027083C:F06
M00003926A:D01	M00004253A:E02	M00005464B:B08	M00027152D:H06
M00003928D:A04	M00003997D:G03	M00005505B:D10	M00027209D:B09
M00003935D:E04	M00003998C:D04	M00005509D:G05	M00027339D:E10
M00003985B:F06	M00004027C:E06	M00005614A:B07	M00027282D:G01
M00004063B:B12	M00004059D:A09	M00005721C:A12	M00023287A:D08
M00004101A:C12	M00004087B:D05	M00005705D:G09	M00026928A:B06
M00004104C:F06	M00004114C:B09	M00005709D:H05	M00027028B:C12
M00004107A:E02	M00004140B:C02	M00004859D:D01	M00027115B:G04
M00004108B:D04	M00004149C:D11	M00005342D:E04	M00027096B:A01

Table 13			
ES55	**ES56**	**ES57**	**ES58**
M00003856A:H10	M00004168D:F05	M00005363D:C05	M00027154B:D05
M00003908C:C04	M00004176B:H09	M00005353C:H01	M00027164A:A09
M00003895C:F05	M00004173A:D03	M00005386C:G01	M00027218C:D06
M00003939B:C02	M00004209B:G01	M00005388B:B02	M00023343B:C08
M00003997A:C08	M00004253D:D04	M00005396C:H04	M00026871C:F12
M00004066D:C02	M00004275A:H07	M00005434A:F11	M00026882A:E07
M00004105C:C05	M00004269C:B10	M00005434C:E02	M00027067B:E09
M00003788B:C08	M00004298A:H09	M00005473C:F02	M00027062C:C04
M00003788C:C05	M00004347A:F10	M00005459B:A01	M00027131C:E07
M00003835B:C05	M00004337A:A07	M00005469A:D10	M00027137D:F05
M00003820B:G04	M00004372A:A08	M00005505D:H08	M00027204B:A08
M00003888C:G08	M00004406D:E11	M00005509B:E10	M00027188A:D12
M00003977D:H04	M00004449B:B05	M00005616B:E11	M00027190B:F06
M00004029D:H03	M00004507A:F11	M00005589B:H12	M00027193A:F07
M00004034A:A05	M00004276A:C06	M00005721D:B03	M00022362D:G11
M00004140D:E03	M00004270C:H05	M00005698A:H12	M00007947B:F07
M00003775C:C01	M00004343A:G07	M00006613C:C02	M00007948B:B07
M00003776B:F08	M00004344B:C06	M00006617A:A06	M00008003B:F09
M00003839D:C03	M00004373D:G10	M00006584D:D01	M00008054C:C03
M00003818C:D02	M00004368A:G11	M00006594B:D05	M00008075D:B01
M00003820C:E08	M00004371B:A05	M00006600D:G07	M00022074A:F05
M00003822A:D02	M00004403A:A02	M00006631D:G09	M00007943C:B02
M00003877C:G01	M00004445D:A04	M00006635A:C01	M00008002B:F09
M00003880A:G10	M00004447A:A10	M00006726D:H10	M00021653C:B06
M00003919D:F01	M00004603D:D09	M00006874D:E01	M00021851D:H06
M00003960D:E09	M00004326D:D06	M00006882C:D03	M00022015D:C11
M00004081A:E11	M00004323B:G12	M00006925B:B02	M00022018B:E09
M00004085B:D12	M00004350A:C04	M00006946B:C08	M00022095C:F03
M00004142C:A06	M00004357A:B10	M00006949B:C07	M00007996C:B11
M00004135D:D01	M00004360B:B08	M00007026A:A03	M00007977B:C11
M00004198B:G08	M00004385D:D06	M00006712A:F01	M00008088D:B01
M00004185B:H03	M00004414D:A01	M00006727A:H12	M00021676B:B12
M00004187A:B05	M00004415A:A01	M00006815D:D11	M00021972A:C10
M00004251B:H12	M00004423A:B05	M00006805D:H12	M00022099C:A10
M00004232D:G11	M00004423C:F03	M00006934B:B11	M00022106D:B06
M00004240A:D03	M00004426B:H06	M00007019B:G01	M00007978B:C04
M00004285C:B06	M00004504C:G07	M00007038D:D01	M00008053D:E09
M00004292A:C08	M00004466A:E04	M00007041C:C05	M00021669B:G02
M00004335A:G05	M00004498D:A11	M00006630A:E05	M00022118A:D08
M00004240C:A06	M00004292A:F03	M00006623C:G07	M00022251A:F07
M00004249A:C09	M00004280D:D10	M00006694D:G06	M00022235D:F07
M00004335D:D03	M00004286D:D02	M00006668D:B10	M00022240C:B03
M00004378A:H10	M00004870D:E05	M00006688A:F09	M00022406C:G03

Table 13			
ES55	**ES56**	**ES57**	**ES58**
M00004381A:E10	M00004871C:C04	M00006745B:C05	M00022459C:G05
M00004444C:H11	M00004872A:D07	M00006846A:B03	M00022627B:D01
M00004225A:E03	M00005395D:D11	M00006823A:H06	M00022184D:F07
M00004284A:C09	M00005395D:B12	M00006925A:B09	M00022177D:G02
M00004264B:F03	M00005412D:G07	M00006894D:A07	M00022460C:E12
M00004404C:B03	M00005413D:G12	M00006895D:A02	M00022627A:A02
M00004410A:F06	M00005513A:H01	M00006991B:E05	M00022144D:D09
M00004412A:G05	M00005515D:G02	M00006994A:C12	M00022203B:A05
M00001340C:A08	M00005607A:C08	M00007046D:E10	M00022214C:C11
M00001340C:D09	M00005366D:E12	M00006577A:B01	M00022252C:A04
M00001395D:B04	M00005618C:H11	M00006630A:E09	M00022420B:C08
M00001466C:H11	M00005708C:D11	M00006619A:G11	M00022640B:G10
M00001528D:B12	M00005810B:C07	M00006704A:C11	M00022641C:H03
M00001517C:A10	M00006795C:B12	M00022127C:E01	M00022652B:G06
M00001561A:G10	M00006755C:C03	M00022128A:C05	M00022216C:H02
M00001565C:F06	M00006756D:G07	M00022176D:F05	M00022199A:F09
M00001569A:H01	M00006779D:F03	M00022214A:H05	M00022214A:D01
M00001341A:H10	M00004821D:C03	M00022220B:B06	M00022273A:B03
M00001375C:C11	M00005358A:H03	M00022278C:E04	M00022256D:G11
M00001397C:F01	M00005480C:A04	M00022282A:A11	M00022261C:D06
M00001431A:F03	M00005481C:H05	M00022260C:H07	M00022490B:G12
M00001457D:E08	M00005490B:B02	M00022263A:C01	M00022648D:G11
M00001505C:C10	M00005820A:H11	M00022377A:E02	M00022709A:G02
M00001615A:D01	M00006621B:B06	M00022399C:B02	M00022701C:A05
M00001618C:E01	M00006752C:D04	M00022056C:D12	M00022826A:C08
M00001358C:D09	M00006757D:H04	M00022087A:D01	M00022963A:E07
M00001360B:B01	M00005000A:H05	M00022088B:E05	M00022904D:D04
M00001391C:B05	M00005296D:G03	M00022090D:B03	M00023095C:A09
M00001389B:B12	M00005378B:B04	M00022094A:A09	M00022684C:C12
M00001485A:C04	M00005461C:D11	M00022096B:D10	M00022765B:E03
M00001559D:E02	M00005464D:D07	M00022176A:F02	M00022898C:H07
M00001545D:F12	M00005657B:F11	M00022217B:E03	M00022902B:F10
M00001549C:F10	M00006596D:H02	M00022259A:D04	M00023003A:H01
M00001579C:E07	M00005826B:F10	M00022381B:C12	M00022768A:A10
M00001630A:E08	M00006577B:F01	M00022399D:A07	M00022834A:H02
M00001386B:E01	M00006582A:F12	M00022401C:G07	M00023002A:C02
M00001389A:F03	M00006664A:C05	M00022407D:G07	M00023003C:C10
M00001418C:F06	M00006678C:B07	M00022417B:C01	M00023012A:C06
M00001454D:H09	M00006840A:A12	M00022435C:C05	M00007973D:B03
M00001442D:D09	M00005020B:D10	M00022471D:A05	M00007939A:F06
M00001450D:H12	M00005296B:H07	M00022464D:F12	M00007941D:D07
M00001479D:B10	M00005403A:D12	M00022469A:A05	M00007948D:F08
M00001598C:F02	M00005376B:E08	M00022500B:D01	M00008012D:H04

Table 13			
ES55	**ES56**	**ES57**	**ES58**
M00001594A:H01	M00005378C:B12	M00022506D:B03	M00008014D:A11
M00001657D:D07	M00005397A:G08	M00022542A:B06	M00008048C:A08
M00003772C:F12	M00005449D:D04	M00022527D:A09	M00008099A:C12
M00003844D:B02	M00005465A:A07	M00022568B:D03	M00021668D:G09
M00003845B:A04	M00005648C:C11	M00022561D:E06	M00021861C:B08
M00003845C:F08	M00006595C:B08	M00022687C:C11	M00021980A:F03
M00003848A:E08	M00006816D:D08	M00022695D:B02	M00007931A:B07
M00003880C:D06	M00006835D:C08	M00022425A:F11	M00007948C:G01
M00001647D:A02	M00006914C:D07	M00022434D:B06	M00007969B:E10
M00001655C:F07	M00007177A:G07	M00022460D:C07	M00008012B:C05
M00003804D:F12	M00006920B:H07	M00022510A:B09	M00008012D:E07
M00003884C:G09	M00007161C:D12	M00022501D:A09	M00008014C:H01
M00003916D:A10	M00006968D:H02	M00022541D:G06	M00008016C:E06
M00003943B:C12	M00006936C:G11	M00022527B:H05	M00008052C:G11
M00003935A:C04	M00006945D:A07	M00022538D:B02	M00008054C:E07
M00003937D:F09	M00007047C:H04	M00022559D:F10	M00008093C:G08
M00001683B:F12	M00007065D:A03	M00022569D:H03	M00021614A:C09
M00001669B:H04	M00007079D:H01	M00022601A:A09	M00008094D:C02
M00003762D:C02	M00006968A:H05	M00022604A:F06	M00021667C:G10
M00003788D:E06	M00007078B:H04	M00022684B:F11	M00021674A:B07
M00003824A:B11	M00007186A:A12	M00022702A:D10	M00021846B:F05
M00003865B:D10	M00004852B:H08	M00022691A:G01	M00021847B:A09
M00003870C:H03	M00005382A:G09	M00022696A:H03	M00021963C:H04
M00003901B:C02	M00005418C:B09	M00022444B:C04	M00007985C:G07
M00003893A:D03	M00005420C:E03	M00022447A:H06	M00008001D:F11
M00003931A:G01	M00005450C:G09	M00022488C:H02	M00007992A:G04
M00003973A:D09	M00005444D:D01	M00022522B:A05	M00008000D:B06
M00001660A:B10	M00005494C:F08	M00022513C:G04	M00008001A:G11
M00003761C:C05	M00005479C:A05	M00022517C:B01	M00008044C:A05
M00003829C:G07	M00005486A:F07	M00022546B:F12	M00008085B:G01
M00003833D:F11	M00005538C:H11	M00022591C:F03	M00008082B:C05
M00003879D:A09	M00005648C:E10	M00022617B:A01	M00008083A:H11
M00003880B:B08	M00005621A:B05	M00022681D:H10	M00021624B:E11
M00003861D:G10	M00004847D:G01	M00022659B:C01	M00021689A:G05
M00003876C:G11	M00005342B:G01	M00022664C:G10	M00021865B:F06
M00003877C:C11	M00005305A:H01	M00022711B:A05	M00021879B:C11
M00003902C:D02	M00026906B:G03	M00022704A:H08	M00021958A:A03
M00003933A:B04	M00026872A:C10	M00022449D:B05	M00021945A:B04
M00003923D:A03	M00026964C:H02	M00022548A:F02	M00021981D:A11
M00003989D:A02	M00026982C:D08	M00022590D:E08	M00007987A:D10
M00003991A:D05	M00027069D:F02	M00022622A:E08	M00007998C:B04
M00004030C:E05	M00027042D:E02	M00022655A:F09	M00008001B:E11
M00004048A:E10	M00027056B:H07	M00022664A:E04	M00008045A:B05

Table 13			
ES55	**ES56**	**ES57**	**ES58**
M00006680D:A01	M00027137C:A03	M00022720A:C01	M00008023A:B03
M00006688C:C12	M00027184D:H02	M00022722D:C07	M00008027D:H09
M00006740A:A06	M00027189C:D04	M00022746D:D05	M00008044B:F07
M00006757A:C09	M00027196A:A10	M00022772A:A06	M00008089C:B08
M00006859D:E11	M00027357D:A02	M00022813C:B09	M00021620D:B06
M00006917B:C05	M00027369A:B03	M00022853D:C05	M00021624B:D03
M00006919A:H12	M00027439B:A09	M00022843A:D02	M00021628C:B09
M00006993B:F02	M00027393D:F01	M00022844C:A01	M00021680D:H08
M00007093C:C11	M00027557D:B06	M00022968D:G06	M00021687C:A04
M00007047D:C02	M00027502C:H02	M00023023B:A05	M00021696C:E02
M00007064B:E09	M00027507C:C06	M00022716A:C01	M00021698A:H03
M00007121A:G04	M00027529B:B11	M00022725D:G05	M00021864C:C07
M00007107C:D02	M00027438D:A03	M00022817D:B09	M00021958A:A04
M00007178D:A10	M00027388A:G05	M00022848D:H09	M00021949D:A05
M00007156D:E11	M00027396C:B06	M00022884D:A07	M00021951B:A01
M00007172D:H03	M00027551C:B07	M00022983A:H04	M00022001B:H10
M00007175D:G02	M00027518B:B07	M00023034B:B10	M00022001D:E06
M00007121D:A11	M00027528A:G03	M00023038D:D04	M00022071D:C08
M00007101C:H01	M00027759B:E11	M00022743C:G05	M00022078B:B04
M00007104D:D10	M00027728A:B03	M00022734C:A03	M00022113B:A12
M00007116A:C08	M00027484A:G03	M00022737D:B02	M00022138C:B07
M00007152A:A10	M00027752B:E05	M00022801A:G04	M00022152A:G05
M00007179B:H04		M00022838B:E05	M00022158C:C08
M00007157B:B04		M00022856A:B09	M00022192B:H07
M00007167C:B10		M00022902C:F11	M00022233C:D11
M00007175B:B11		M00022893D:C06	M00022252A:C01
M00007177B:C02		M00022922D:G06	M00022370A:G07
M00007141A:G08		M00022986B:C02	M00022300A:A05
M00007196D:D02		M00023002D:C12	M00022386D:C04
M00007145C:B05		M00023096C:A03	M00022072D:E12
M00007126D:H01		M00023097A:C03	M00022102D:A10
M00007140C:G12		M00022743C:G06	M00022207C:C01
M00007200A:B12		M00022736B:B03	M00022249C:G09
M00007203C:E06		M00022737B:F12	M00022383C:F05
		M00022831C:F11	M00022384B:E06
		M00022836C:A07	M00022067A:B03
		M00022854D:C04	M00022056B:G12
		M00022860A:A07	M00022084B:C03
		M00022861C:B04	M00022087D:F12
		M00023096A:F03	
		M00023096D:B11	
		M00023097C:D10	

Table 14

ES59	ES60	ES61	ES62
M00001418A:A02	M00001477A:G02	M00004450A:G07	M00005515B:B08
M00003877C:A08	M00003853C:A09	M00004353D:C06	M00005385B:A10
M00003977C:D01	M00001694B:H12	M00004406A:H12	M00005516D:F12
M00004295A:C02	M00001664D:E02	M00004048C:C02	M00005822D:C05
M00001383C:C04	M00003847B:H01	M00004170B:G04	M00004841C:H03
M00001500A:A02	M00001631D:G08	M00004108C:D07	M00005810B:G02
M00003880B:D03	M00004498D:F02	M00004125B:A02	M00007107A:H08
M00003803B:G12	M00001563A:F04	M00004109A:B07	M00004825A:G12
M00003819D:B02	M00001558D:E02	M00004123B:G05	M00005327C:G08
M00004178B:F07	M00004278C:H11	M00004152A:F03	M00005390C:E05

ES63	ES64	ES65	ES66
M00005520A:H11	M00006790D:F10	M00027175D:A05	M00026949A:F04
M00006814D:D09	M00006627C:C02	M00026910C:C05	M00023432D:F09
M00006918D:G08	M00027462D:A12	M00027280D:H01	M00027178B:E04
M00007197D:D12	M00026972A:F04	M00023289D:E06	M00027225B:D03
M00005497C:G08	M00027592D:C05	M00023373A:D01	M00023340B:B07
M00007109D:G01	M00026945B:C10	M00027231A:D01	M00027283C:H12
M00005377C:F07	M00027231C:D08	M00023321A:F07	M00027085C:H12
M00006813B:E04	M00027083D:F06	M00027266C:G12	M00027234C:B05
M00005825A:A10	M00027142A:C01	M00023398D:F10	M00023390A:C04
M00005416B:A01	M00027607A:A09	M00027603C:E02	M00026810A:H04

ES67	ES68	ES69	ES70
M00023340B:H12	M00027642C:D11	M00022714B:D04	M00022709A:C01
M00027237C:D04	M00027202B:B09	M00022838A:H05	M00022413B:D07
M00026809C:D10	M00027459A:G12	M00022392C:H06	M00022467C:H07
M00027386D:C02	M00027250A:C04	M00022363C:D03	M00022561B:B09
M00027343B:H05	M00027499B:G02	M00022205A:C02	M00022214C:E09
M00027356A:H02	M00027053C:B06	M00022717C:F05	M00022697A:C08
M00027363D:A08	M00027598C:D06	M00008015B:D08	M00022682A:F10
M00027364D:E08	M00006989C:B01	M00021625B:G07	M00021841A:E11
M00027618A:B08	M00006837B:H12	M00008100D:C08	M00021691B:E04
M00027628D:D08	M00007202A:A09	M00022669D:G07	M00022477C:C07

Table 14

ES71	ES72	ES73	ES74
M00022134D:D12	M00008028D:B01	M00022513C:E10	M00023363C:A04
M00022705B:F08	M00021931B:F04	M00022518C:C04	M00001401B:A02
M00022903D:H02	M00008097C:E04	M00022544C:D08	M00008023C:A06
M00022915C:C09	M00008082B:H10	M00022785C:B10	M00022077D:A12
M00007965C:B02	M00008006A:H02	M00022525C:E09	M00023284B:G06
M00022368C:C11	M00022167B:H02	M00022641D:F08	M00023369D:C05
M00007937C:E08	M00022509D:A12	M00022923A:A09	M00023413D:F04
M00021852C:D12	M00022169A:E11		M00026905A:G11
M00008000D:G11	M00022184D:H07		M00027169D:H06
M00021908B:F03	M00022441B:A06		M00005434D:H02